The Design and Implementation of Workflows for Social Simulation

Mark Birkin, Nick Malleson
School of Geography
University of Leeds, UK

Andy Hudson-Smith
Centre for Advanced Spatial Analysis
University College London, UK

R.J.Allan
STFC Daresbury Laboratory
Warrington, UK

Meik Poschen, Rob Procter
Manchester eResearch Centre
University of Manchester, UK

Sean Bechhofer, Alexandra Nenadic, Carole Goble
School of Computer Science
University of Manchester, UK

Richard O. Sinnott
Melbourne eResearch Group
University of Melbourne, Australia

Abstract—**This paper reports on progress towards the development of an e-research infrastructure for social simulations. We argue that simulation models are increasingly of interest to social scientists across a wide range of disciplines, and of value to both academics and policy-makers. The constituent services of an e-infrastructure for social simulation are described, and examples of social simulation workflows that connect these services into more substantial general purpose tasks shown. The paper discusses the ways in which workflows might themselves be embedded into portals for both spatial analysis and decision support, and reviews some of the outstanding barriers to further development.**

Keywords: Workflows; Social Simulation; Modelling; Web Services; Taverna

I. Background to Social Simulation

The technique of numerical simulation is well established in the social sciences. Cellular automata models can be traced back at least to the 1960s with Schelling's elegant demonstration of the emergence of segregation in an urban housing market from simple, weak preferences for social harmony [1]. Related approaches in microsimulation can be found even earlier in the 1950s in the work of Orcutt in economics and Hagerstrand in human geography (see [2] for a more complete review). More aggregate approaches to modelling social interactions were already widespread by the 1960s but can be traced right back into the nineteenth century [3].

Recent interest has swelled following dissemination of agent-based approaches [4]. Such methods provide a flexible means of representing theories of social behaviour, in which restrictive assumptions about normative behaviour can be steadily relaxed and tested or calibrated with social data which is increasingly ubiquitous. Examples can be found across the social sciences, for example in the formation of economic markets [5], the outcomes of political or group negotiation [6], health care policy [7] and the behavioural interaction of both criminals and victims [8]. In many cases multi-agent models are seen as theoretical frameworks in which rich systemic behaviours may be found to have arisen from sparsely identified processes [9], but others have argued for applications with much more obvious policy-relevance [10,11].

Elsewhere, we have argued that the time is right for the promotion of social simulation approaches to a more universal community of developers, users and analysts, through their deployment via a research e-infrastructure [12]. Such infrastructure might provide sharing of techniques, powerful visualisation of model outcomes, access to high performance computational resources and access not only to widespread secondary data sources but increasingly the means for primary, crowd-sourced data capture [13].

In this paper, we describe a proposed architecture for e-research in social simulation based upon the JISC-funded National e-Infrastructure for Social Simulation (NeISS – www.neiss.org.uk) project. We present the range of models that can be supported by such an architecture and the range of problems and scenarios that may thus be underpinned. We discuss the various component services that need to be mobilised in order to enable such an e-infrastructure and describe how we have begun to bring together these component services into social simulation workflows that combine data into models and subsequently allow the visualisation and publication of outputs. We seek to illustrate how workflows will begin to provide essential building blocks through which the value of the e-infrastructure may be realised. We present some real examples and begin to fix ideas about the ways in which social simulation workflows might be used in a practical context. Finally the paper concludes with a discussion of the key issues that will inform our plans for the immediate future.

II. ARCHITECTURE AND SERVICE COMPONENTS FOR SOCIAL SIMULATION

In an earlier project, we explored the use of computational services on an e-science grid to address the problem of finding ideal locations for the provision of health care services. This is a useful and representative problem. Data on patient demographics, accessibility to services, the economics of provision, and most likely existing patterns of utilisation are all required. Appropriate models allow the impact of alternative networks of provision to be evaluated and performance to be optimised. Visualisation of the model outcomes has obvious value in the analysis and interpretation of model outcomes.

The exploitation of grid infrastructure has traditionally been used for the execution of computationally expensive simulations but increasingly grids and their evolution have been applied to establish collaborative problem solving environments [14]. The problem solving environment we aim to support in NeISS needs to combine behaviour modelling, 'what if?' impact analysis and location optimisation for value planners and policy makers with a responsibility for efficient resource allocation, as well as to academics trying to understand the structure and evolution of complex social systems. In related work, it has been shown that the same framework is equally applicable to a wide range of problem domains, such as provision of housing, transport, retailing and education services.

The e-infrastructure and component services of the NeISS project are shown in Figure 1. In broad terms, the overall objective of the e-infrastructure is to support the social simulation 'lifecycle' in either research or policy applications. The lifecycle has four essential components: data, models, interpretation and publication. These will be addressed in the list of services in Table 1.

In the past, social simulation has tended to rely on either secondary data sources [15] or on abstract idealised representations with synthetic data [9]. The NeISS architecture recognises that primary data sources, especially those that are either crowd-sourced or sensed, will be of increasing importance in the future [16].

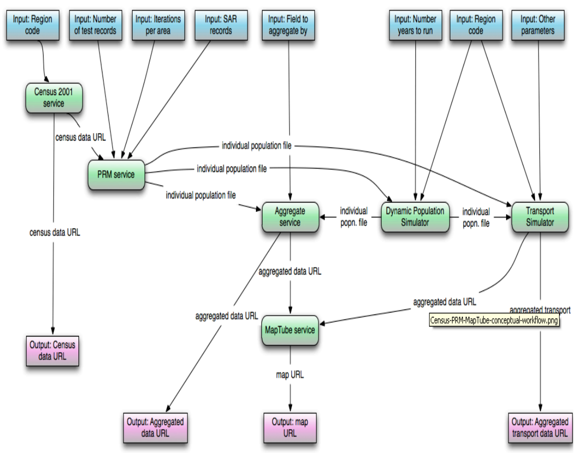

Figure 1. Current NeISS components and services and their interactions

TABLE I. CURRENT NEISS SERVICE PORTFOLIO

Name	Type	Description
Census2001	Data	Secondary data service
PRM	Model	Population reconstruction model
DSM	Model	Dynamic simulation model
Aggregator	Data	Data aggregation service
SurveyMapper	Data	Primary data service
TSM	Model	Transport simulation model
FusionTool	Data	Data fusion service
MapTube	Interpretation	Map creation service

With regards to models, the NeISS incorporates three specific model groups:

- Population Reconstruction Model, which allows for the creation of synthetic micro-populations for a city or region;
- Dynamic Simulation Model, which projects the population of a city or region forward in time;
- a suite of Spatial Interaction Models, which reproduce the behaviour patterns by which diverse urban services are patronised.

One of the key aspects in the interpretation of model and simulation outputs is the use of maps for the diagnosis of spatial trends and patterns. Statistical analysis and decision graphics (charts, reports, tables) are also important elements in this stage of the lifecycle.

Publication of simulation processes or outputs supports data intensive (social) science by providing a platform through which simulation experiments can be documented, shared, updated and extended. One of the most important ways in which a particular experiment might be published and shared is as a workflow. Workflows allow for the definition of the interactions and data flows between services, typically to capture a scientific process. Table 1 provides a description of the services that are currently available or under development in the NeISS environment. This combination of services provides extensive coverage of the requirements of the social simulation lifecycle.

The Census2001 service provides a means to exploit the considerable resources of the UK Census, providing access to key data in six tables for over 220,000 neighbourhoods ("Output Areas"). The SurveyMapper service provides the means to generate primary data for specific projects or applications, and allows users to design and distribute web questionnaires on a subject of their choice. The technology has been used to map attitudes to the UK 'credit crunch', and to local traffic taxes and congestion charges, both commissioned and reported by the BBC. The Aggregator service provides a means to link together individual data from disparate sources using common characteristics. For example, this could be used to combine survey data with demographic data, and thus to present reweighted and localised generalisations of social survey data such as the British Household Panel Survey (BHPS – http://www.iser.essex.ac.uk/survey/bhps) as well as primary crowd-sourced data. The FusionTool is a refined version of the Aggregator service that allows flexible matching between fuzzy data entities. The Population Reconstruction Model combines small area census data with samples of anonymised records in order to create synthetic and household populations for a city or region. The Dynamic Simulation Model uses representations of key demographic processes – births, deaths, migration, marriage, household formation and dissolution – to project local populations forward through time. The projection model runs 30 years into the future from a base period in 2001. The Transport Simulation Model (TSM) is an instance of spatial interaction modelling which assigns traffic behaviour (mode and route choice) on the basis of residential location and activity patterns in relation to fixed transport infrastructure. By varying the infrastructure, the impact of alternative transport policy scenarios can be simulated. The MapTube service provides a means for visualising spatial distributions from any of the data or model services in NeISS, overlaying maps and boundary data from GIS shapefiles or third party sources such as GoogleMaps.

III. WORKFLOWS FOR SOCIAL SIMULATION

A. Overview

Scientific workflows [25, 26] have emerged to tackle the problem of excessive complexity in scientific experiments. They provide a high-level declarative way of specifying what a particular scientific process modelled by a workflow is set to achieve and not how it will be executed.

Each workflow component is responsible for a small fragment of functionality; many components (typically Web services) need to be chained in a pipeline in order to obtain a workflow that can perform a useful task. A further advantage of using workflows is the potential to automate highly repetitive processing that research work is known to involve. Finally, workflows promote collaboration and support reproducible science through sharing and reuse [27].

Scientific workflow management systems, such as Taverna [28], provide a mechanism to automatically orchestrate an execution of services, coordinating processes and managing the flow of data between them. Taverna workflows can easily be shared through myExperiment [29], a social networking and workflow sharing environment for scientists.

B. The NeISS Workflow

In this section of the paper we will describe the NeISS workflow, which has been constructed from the services described previously. The workflow begins with access to the Census 2001 Service, for which the key input parameters are the selection of a city or region of interest and a collection of variables associated with that area. This produces a results spreadsheet file of census data for the given region. The URL to the census data file is then passed to the PRM service. The parameters for this service are assigned to trade off between simulation accuracy and speed of execution.

The PRM service itself produces a population of individuals that is subsequently passed to the Aggregator service. The Aggregator links individuals in the simulated population to individual level data resources such as the BHPS. The service requires the selection of a field or variable to aggregate by. Once the individual population has been combined with BHPS data it can be re-aggregated and mapped. The service produces a data resource with the aggregated data at output area level.

Prior to the enactment of the Aggregator, it is also possible to invoke a Dynamic Simulation Model (DSM) which rolls the population forwards in time, including the merger, deletion and addition of new population and household records. The DSM is actually nine separate modules wrapped into a single service within this workflow. It produces a file of data aggregated by geographical areas.

Like the DSM, the Transport Simulator (TSM) is also a bundle of software modules wrapped as a single service. Key parameters for this process include switches for scenarios such as new roads or congestion charging for city centres. The TSM generates a file of important environmental and economic performance indicators (such as air quality and journey cost) that can subsequently be passed on to the MapTube service.

Finally, the URL of the aggregated data file from either the Aggregator or the TSM is visualised using the MapTube service. MapTube uses default settings for visualising the map and produces a URL at which the map can be seen.

C. Workflow Enactments and Security

The NeISS services described above can be accessed through a Liferay portal (operational) or eventually through a Sakai portal (currently under development). The NeISS services can either be accessed individually, through specific "service portlets" (e.g. the Census2001 service portlet, the PRM service portlet), or through a number of pre-defined workflows that chain the NeISS services together and are exposed through "workflow portlets". In the workflow portlets, users will be able to specify input data and other parameters and submit workflows for execution on the Taverna Server. Once the execution has finished, users will be able to collect the results from the Portal.

Several security issues need to be addressed in the proposed architecture. The first one is how to authenticate the user to the Portal and determine what the user is allowed to see and do in the Portal, e.g. what workflows is the user allowed to run and with what data. The second one refers to workflows that need to access secure NeISS services and how the Taverna Server is to invoke such services during workflow execution on behalf of the Portal user (the Taverna Server acting here as a proxy). Finally, we need to solve the issue of passing the selected authenticated Portal user's credentials (required to invoke secure NeISS remote services) to the Taverna Server, so it can enact the workflow containing such services.

Access to the NeISS Portal is through Shibboleth and the UK Access Management Federation [30]. The Liferay portal content (its portlets) is configured using sofware developed in the OMII SPAM-GP project, in particular the Content Configuration Portlet (CPP) [31]. This allows the security-oriented information (roles/ licenses) released from the Shibboleth Identity Provider (IdP) to be extracted from the eduPersonEntitlement attribute and used by CPP. A similar mechanism is being investigated for Sakai, which has an internal role-based security structure.

Data sets required by NeISS researchers may have strict access and usage conditions, thus data services, such as the Census2001 service, need to be protected and require authorisation information to be provided. The current authorisation model, which also applies to the enactment of the workflows, is based upon a hybrid-oriented model as outlined in [32]. Here the user, via the IdP, pushes the roles and attributes that are needed for future authorisation decisions to the Portal. These are kept and stored in a local LDAP server, which is one of the portal's back-end services. Roles and attributes can also be pulled from other known and trusted attribute authorities. More precisely the Policy Enforcement Points (PEP) for the portal services are configured to recognise particular sources of authority.

It is the case, however, that access to Grid based resources using X.509 credentials is not currently aligned with the SAML assertions issued by the UK Federation. This poses a problem if large social simulations need to be run on Grid resources. There are several approaches to overcome this. The "traditional" method is to download a proxy credential from a MyProxy service available through the Portal where the user already has a credential stored. In this case the user has to be aware of the use of X.509 certificates. Alternatively it is possible to use additional Shibboleth information such as the encrypted MyProxy username and passwords of users to automatically create the proxy credential as part of building up the portal session. A further solution that has been taken by the SARoNGS project [35] is the automatic creation of low assurance proxy credentials directly from SAML assertions via the Credential Translation Service.

In the proposed architecture, the Taverna Server is configured as a portal back-end service alongside LDAP and a database, such as MySQL or Oracle, which are used by all the portal applications (tools) exposed as portlets. Access to the Server will be restricted to the Portal only.

At the moment, secure NeISS services use various authentication models. The Census2001 service uses Globus GSI-based authentication with client proxy certificates, the MapTube service requires username and password with HTTP Basic Authentication and some services do not employ security at the moment but are planning to, etc. Users need to provide each of the credentials required by the services in the workflow if the overall enactment is to be successful. We are leaning towards the unified approach where all NeISS services will be using the same security model based on grid proxy client certificates that users will obtain once they authenticate to the Portal. This way, only one credential will be required to access all secure NeISS services. The issues of security-oriented workflows and their realisation are also described in [33, 34].

In the following we describe the proposed interactions between the Portal and the Taverna Server (depicted in Figure 2):

1. The workflow and its inputs are uploaded to the Taverna Server through a targeted portle using the Server's REST API. Depending on the tool permissions (set by user roles) they will be available only to a particular user portal session or shared among a group of users. In the

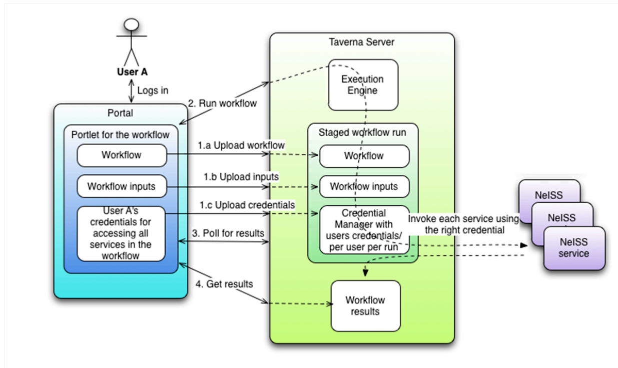

Figure 2. Portal and Taverna Server

case of Sakai this can be managed using the Resource folders, which have an internal service API making them a repository for other tools to use.
2. Additional user attributes including roles and licenses for access to external services can be automatically pulled from the LDAP back end service or other authority as required, e.g. for access to data in the workflow. User's credentials required to access secure services in the workflow are gathered in the Portal, packaged, encrypted and pushed to the Taverna Server. Ideally, only one credential (i.e. the Globus proxy client certificate) will be required to access all services. After this set-up stage (uploading the workflow, inputs and credentials), the actual workflow execution can be requested from the Taverna Server.
3. The workflow will effectively run independently of the Portal once invoked. A status portlet is therefore provided that will periodically poll the Taverna Server for results.
4. Once the workflow has completed, the results are ready and can be viewed or downloaded using the status portlet.
5. The complete history of the workflow can be saved (e.g. in the Resource folders) to be shared and re-run on the Taverna Server or it can be deleted. Information can then also be "published", e.g. to myExperiment or MapTube.

Work is currently on-going to add Globus GSI-based security to the Taverna Server, which will enable it to authenticate the user, on whose behalf it is enacting the workflow, to the remote NeISS service via X.509 client grid certificate.

There are a number of issues raised by this work. Firstly identity management is required with mapping among a number of unique but different user credentials such as internal portal id, Shibboleth id, X.509 DN, etc. The appropriate identities must be selected to access services contained in the workflow via the Taverna Server. These services themselves employ a variety of security models and may access different attribute authorities. This introduces difficulties when enacting a complex workflow. In some cases, user's credentials must be provided to external services via the Taverna Server, but a delegation mechanism is not always available, This raises additional security concerns. Finally, should a user be able to begin to enact a workflow if they do not have sufficient privileges to enact every service in that workflow? Is it possible to know in advance if all the services will authorise the user? Sharing of secure workflows by users adds to these numerous challenges. Many of these and related issues are described in [31].

IV. CURRENT APPLICATIONS

In this section of the paper we discuss some indicative applications of the infrastructure. Our main purpose here is to illustrate and discuss the features of the infrastructure, rather than the substantive outcomes and potential policy implications of the models and analytical components.

Figure 3 reflects the combination of services Census-PRM-Maptube. These basic components were integrated in the first workflow generated by the NeISS project. In this process data is extracted from a suite of census tables, then used to create a synthetic population of individuals and households with basic demographic and housing attributes (age, gender, marital status, ethnicity, quality of health, social group, housing type, tenure). Any of these characteristics can then be mapped spatially and passed to the MapTube service for visualisation. In this case, the population aged 65 and over shows a characteristic 'doughnut' effect, with a heavy concentration in the outlying peri-urban and rural areas, and secondary concentrations in the central areas. The distributions generated through this process are not new data, indeed the PRM is designed to reproduce small area characteristics as closely as possible, hence the same map could have been produced by putting 2001 census data through a standard GIS engine. It is worth making the point however that to do this a user would need access to the data itself, some means of extraction, and access to the necessary GIS software (and spatial data), probably under licence. NeISS brings these elements together within a single infrastructure. MapTube alone, which has been developed historically as an independent service, at the last count had over 8,000 registered users and more than 520,000 hits on the MapTube website.

For the example illustrated by Figure 4 we have added the aggregator to the service mix. The aggregator ties the synthetic population generated by the PRM to an external database of

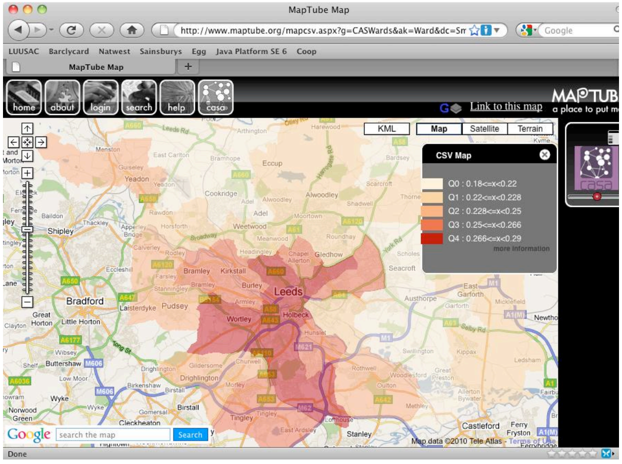

Figure 3. Linking BHPS data to a synthetic population to estimate the number of smokers in Leeds.

individuals and households. In the current infrastructure, the external database is the British Household Panel Survey (BHPS). In future, we hope to extend this feature to include other sources for specific domain applications, such as the Health Survey for England or National Travel Survey. These surveys are typically very rich in the representation of both characteristics and behaviour, but quite restricted in their spatial detail. Each synthetically generated member of the population in a city, region or neighbourhood can now be linked to a member of the external database by matching their demographic characteristics. From this process, additional characteristics are assigned to the synthetic database using the profiles in the BHPS. The power of this lies in the fact that while the PRM generates in the order of 10 demographic and household attributes, the BHPS has over a thousand attribute fields. In this way, many spatial patterns can be newly inferred, for example the distribution of smokers as shown in this illustration (for a discussion of the importance and policy implications of spatial variations in smoking behaviour, see [18]).

Figure 5 demonstrates the use of the dynamic simulation model (DSM). The dynamics are generated through a series of sub-models representing key demographic transitions (i.e. marriage, migration, procreation and so forth – see [19] for more discussion). In this way the population is stepped forwards in time, typically on an annual basis. Some results of running the models forward in time over 30 yearly cycles are shown in this illustration.

The introduction of behavioural simulators is where much of the substantive excitement begins for social modellers and policy-makers; an example of which is illustrated by Figure 6. Most of the urban areas in the UK will experience a combination of population growth, ageing, and ethnic diversification across all planning and projection horizons, from next year to the next fifty years [20]. This will have implications for the provision of health care, education, policing, transport, housing and land use, economic development, and many other activities and services. Our efforts to date have focused on trying to crystallise an

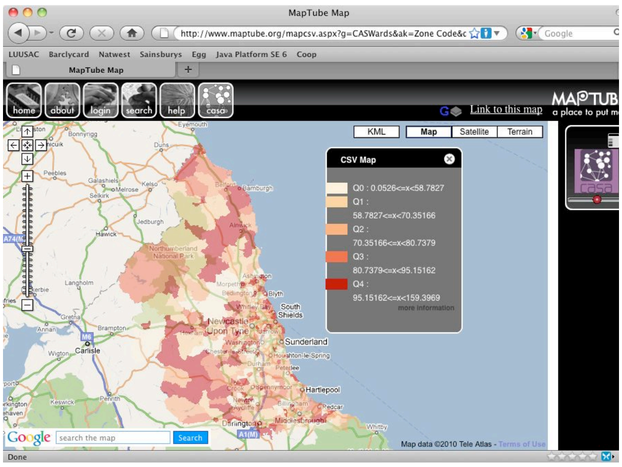

Figure 4. The distribution of elderly people around Newcastle

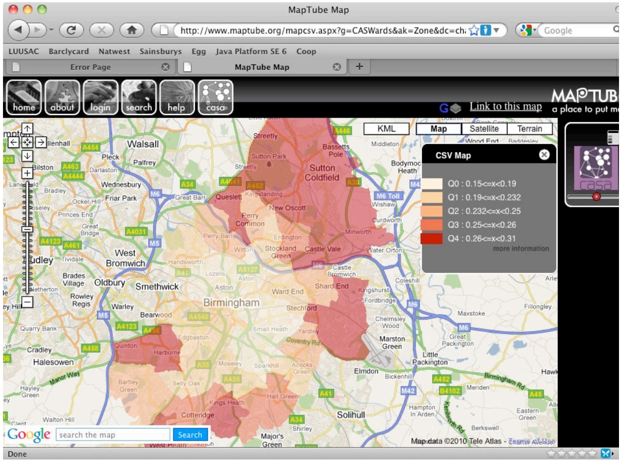

Figure 6. A dynamic simulation showing proportional population change between 2001 and 2031 in Birmingham.

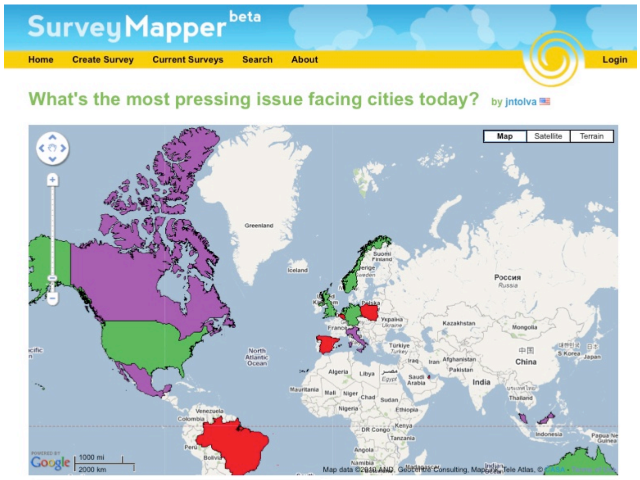

Figure 5. An example of a Survey Mapper global survey.

understanding of capability through demonstration scenarios, and we have chosen transport as the key theme for this purpose. The traffic simulation model (TSM) constructs an idealised representation of roads and public transport networks in a city, and it assigns trips to these networks on the basis of both residential and employment locations. The TSM can then explore the effect of 'what if?' scenario changes in transport provision at any time period. Appropriate performance indicators are generated to support further analysis and interpretation. In this illustration, the TSM is applied to the base period (2001), a mid-term projection period (2021), and to three projection scenarios – congestion charging, road pricing, and promotion of non-vehicular alternatives (see, for example, [21]). Due to limitations of space, these scenarios are presented in terms of aggregate city-wide outcomes. Although road pricing for example appears as a marginally superior option to congestion charges on most indicators, the spatial impacts of these policies may be very different, as perhaps congestion will be most marked in certain areas. This could be tackled most directly through the charge. These distributions can be explored much more effectively through online 'visual surfing' of model options and outputs, so another of the great virtues of the e-infrastructure is that it provides users with the capability to evaluate and interrogate many different policy options and alternative futures.

The last example we use is that of SurveyMapper, illustrated by Figure 7. This model aims to exploit the power of online polling within a social simulation context, and allows users to design and implement their own surveys. At the moment this service is weakly connected to the e-infrastructure, but it does have MapTube embedded as a secondary service so that poll results can be displayed geographically alongside other social or behavioural analyses. This example shows results from a global survey of urban problems. Our longer intentions are to add SurveyMapper to the workflows previously discussed, so that for example if a policy-maker is considering a scenario in which people are to be offered financial incentives to leave a car at home and travel to work by bicycle, then the scheme could be floated in advance to the general public via SurveyMapper. Not only could these responses be used as a basis for calibration of the NeISS simulation tools, but respondents and the public in general may equally be encouraged to assess the impact of these reactions through the various modelling and analysis tools provided by the infrastructure.

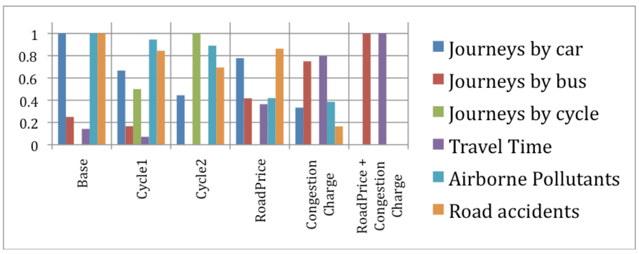

Figure 7 Traffic simulation 'what-if?' scenarios

V. Discussion and Conclusions

The project team has made considerable progress with the definition of a set of services, and with their integration into workflows which facilitate the combination of processes from data extraction, analysis and integration, modelling, simulation and visualisation. The technology is applicable to a wide variety of substantive problems in social science and public policy.

We noted that core functions such as the census data portlet and MapTube visualisation service offer capabilities that are already available by other means, for example within proprietary GIS systems. By making these technologies freely available within an e-infrastructure they can become more widely used without the limitations of expensive licences or restrictive covenants on the secondary use of data. These developments seem well-aligned with current government policies to make data from organisations such as Ordnance Survey and Office for National Statistics much more widely available.

One of the features of MapTube which is not exploited by the current architecture is the ability to archive and share maps. NeISS is working with developers in the MyExperiment group at the Universities of Southampton and Manchester to develop electronic publication and retention capabilities for social simulation, and this will shortly allow mechanisms such as publication of NeISS experiments to social networking sites, and shared access via portable telecommunications devices. We are also exploring different ways of visualising numerical experiments and simulation outputs through the Exhibition Space concept in which the products of services or workflows are ported to a virtual environment in which to be explored and absorbed by users of the infrastructure.

The notion of linkage between datasets via the aggregator function has been touched upon briefly. Through our interactions with the National Centre for e-Social Science DAMES project, we are working to develop a more rigorous approach to data integration. The DAMES toolkit could provide alternative statistical mechanisms for matching and merging between datasets, as well as formalised libraries to add consistency to this process. A more general problem is how a user with specific interests might bring an entirely new dataset into the NeISS infrastructure – let's say the British Crime Survey, or even a newly commissioned data product. It is our intention to increase the range of datasets which are available or accessible to NeISS and in this way to develop further our understanding of the kinds of standards, protocols and metadata descriptions which might be required to support this process. One of the concerns which sometimes arises in work which seeks to establish links between individual data records regards privacy and confidentiality. This seems unlikely to be a practical concern with datasets which have been so heavily synthesised, although further protective strategies such as over-imputation might be considered.

The idea of external links applies not just to data but to a whole variety of tools for modelling, analysis, simulation and visualisation. In a sense the development of SurveyMapper is an example of this, where at present we have effectively a standalone piece of technology and the next challenge is how to integrate this with the existing infrastructure. If this could be achieved as a generic rather than a one-off customised process then the long-term potential of NeISS as an inclusive infrastructure for social simulation and even social science more generally would be greatly enhanced.

It seems likely that an important next step will be to begin to assemble the NeISS services and workflows into some form of decision support system. Similar steps have already been undertaken with the dynamic models as part of the MoSeS and GENeSIS projects [15] and the services and workflows we have described here can all be delivered already through a Liferay portal, which could provide an appealing environment for decision-support applications. The service components themselves are currently distributed between servers in Glasgow, Daresbury, Leeds and London, and we fully intend to maintain this federated structure. Most of the models described here are not enormously onerous, although to run the dynamic projections at higher levels of resolution can take several minutes or hours for a single city region. In the past, grid services have been used for both data storage and processing [22] although possibilities for migration into the cloud should not be discounted in the longer term.

At present, although the infrastructure is still not fully populated, the ability to generate projections and scenarios already outstrips the capacity of the NeISS user community to process and exploit. However, the barriers to sharing amongst the community are clearly significant, some of which are social and others technical. Taking the former first, it is clear from our own previous studies of researchers' attitudes towards sharing that many are reluctant to embrace the vision of the Open Science movement [23]. In general, researchers' attitudes can be summarised as cautiously welcoming the benefits of wider sharing but conscious of the risks of disclosing their findings too early: they want to share but desire to retain control of whom they share with and the timing of publication. As part of the requirements gathering for the NeISS project, we have interviewed a range of potential users, and their views in regard to sharing and re-use of simulations are broadly similar.

Most interviewees agree that a portal/repository of simulation resources would be very beneficial: "*I think it generates the feel of a buzz, like you are in a shared kind of lab, which can be quite exciting – and it can also lead to unexpected conversations*." However, for some of our interviewees sharing raises IPR issues: "*I would be concerned with Intellectual Property issues with this sort of shared information, because these models take ages to build. I don't know that I'd be willing to deposit anything there, personally, if I'm honest.*" Related to this was a more general concern about attribution and credit for contributing would work.

Interviewees also expressed a number of concerns about barriers to re-using simulation resources, and to issues of trust in particular. One of these relates to how the quality of objects offered to the repository would be assured if they had not been subject to some form of peer review. At the same time sufficient documentation has to be provided for the user to be able to understand and validate outputs of a given model or procedure. There is another concern regarding the challenges of understanding a complex computational artefact such as a simulation, so that a potential re-user will be able to decide if it is relevant to their needs and to trust its behaviour and results: "*And I don't know how it feels like using something someone else has created, because I'd like to know exactly how it's done, so I have complete confidence in what I'm doing."* This question is one that we will be attempting to address through the provision of curation tools, procedures and standards.

In summary, there are many challenges to developing complex forecasting tools such as simulations that a diverse range of end-users will be able to use effectively. A key part of the solution, however, is to create the conditions for strong user engagement [24]. The objective of NeISS is to have end-users such as policy-makers, planners and private citizens participating, not just by contributing data through tools such as SurveyMapper, but as consumers, evaluators and as producers of social simulation experiments. In this way, we aim to move towards a "pull model" of e-Science adoption, in which "the user community determines its own information needs and requests (or pulls) appropriate information from the technical experts" [24] and is able to take the lead in commissioning simulations that serve their specific requirements.

ACKNOWLEDGEMENTS

This research has been funded by JISC through the Information Environments Programme. Census output is Crown copyright and is reproduced with the permission of the Controller of HMSO and the Queen's Printer for Scotland.

REFERENCES

[1] Schelling, T. C. (1969) Models of Segregation, American Economic Review, 59(2), 488-93.

[2] Birkin M., Clarke M. (2010) Spatial Microsimulation Models: A Review and a Glimpse into the Future, in Stillwell J, Clarke M (eds) *Understanding Population Trends and Processes*, (Volume 4, Chapter 9), Springer, Berlin.

[3] Ravenstein, E. (1885) The laws of migration. *Journal of the Royal. Statistical Society 48*:167–227.

[4] Epstein, J. (2007) *Generative Social Science: Studies in Agent-based Computational Modelling*, Princeton University Press, Princeton.

[5] Arthur, W. (1999) Complexity and the Economy, *Science*, 284, 107-109.

[6] Axelrod, R. (1997) The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration, Princeton, New Jersey: Princeton University Press

[7] Christiansen J., Campbell A. (2008) HealthSim: An Agent-based Model for Simulating Health Care Delivery, Argonne National Laboratory.

[8] Malleson, N., Evans, A. J., & Jenkins, T. (2009) An Agent-Based Model of Burglary in Leeds, *Environment and Planning B*, 39, 1103-1123.

[9] Epstein, J. and Axtell, R. (1996) *Complex Social Systems*, Brookings Institute, Washington DC.

[10] Ferguson, N. M., Cummings, A. T., Cauchemez, S., Fraser, C., Riley, S., Meeyai, A., Iamsirithaworn, S. & Burke, D. S. (2005) Strategies for containing an emerging influenza pandemic in Southeast Asia. *Nature* 437, 209–214. (doi:10.1038/nature04017)

[11] Jordan, R., Birkin, M., Evans, A. (2010) Agent-based Simulation Modelling of Housing Choice and Urban Regeneration Policy, in Bosse T, Geller A, Jonker C. (eds) *Eleventh International Workshop on Multi-Agent-Based Simulation*, , 145-157, AAMAS, Toronto.

[12] Birkin M, Procter R, Allan R, Bechhofer S, Buchan I, Goble C, Hudson-Smith A, Lambert P, de Roure D, Sinnott R (2010) The Elements of a Computational Infrastructure for Social Simulation, *Philosophical Transactions of the Royal Society A*, forthcoming.

[13] Bell, G., Hey, T. & Szalay, A. (2009) Beyond the data deluge. *Science* 323, 1297–1298. (doi:10.1126/science.1170411)

[14] 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.

[15] Birkin M, Townend P, Turner A, Wu B, Xu J (2009) MoSeS: A Grid-enabled spatial decision support system, *Social Science Computing Review*.

[16] Savage M., Burrows R. (2007) The Coming Crisis of Empirical Sociology, S*ociology*, 41, 885-899.

[17] Hudson-Smith, A., Batty, M., Crooks, A., & Milton, R. (2009) Mapping for the Masses: Accessing Web 2.0 Through Crowdsourcing, *Social Science Computer Review*, 27, 524-538.

[18] Tomintz M., Clarke G.P., Rigby J. (2008) The geography of smoking in Leeds: estimating individual smoking rates and the implications for the location for stop smoking services, *Area* 40(3), 341-353

[19] Wu B, Birkin M, Rees P (2010) A dynamic MSM with agent elements for spatial demographic forecasting, *Social Science Computing Review*, forthcoming.

[20] UK's ethnic minority numbers 'to rise to 20% by 2051', BBC News, 13th July 2010, www.bbc.co.uk, accessed 23/7/10.

[21] Tax-free bikes for work, www.cyclescheme.co.uk, accessed 23/7/10.

[22] 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.

[23] Procter, R., Williams, R., Stewart, J., Poschen, M., Snee, H., Voss, A. and Asgari-Targhi, M. (2010). Adoption and Use of Web 2.0 in Scholarly Communications. In *Philosophical Transactions of the Royal Society A*, Special Issue on e-Science. In press.

[24] Kerr, A. and Reay, D. (2010). Push versus pull models: Engaging business and policy users of e-Science outputs. *AHM*, submitted.

[25] Goble, C., De Roure, D. (2009) The impact of workflow tools on data-centric research. In Tony Hey and Stewart Tansley and Kristin Tolle (eds.), *The Fourth Paradigm: Data-Intensive Scientific Discovery* pp. 137-145, Microsoft Research, 2009. Available at http://research.microsoft.com/en-us/collaboration/fourthparadigm/.

[26] Wolstencroft, K., Fisher, P., De Roure, D., Goble, C. Scientific Workflows, Connexions, November 2009. Available at http://cnx.org/content/m32861/1.3/.

[27] De Roure, D., Goble, C., Stevens, R. (2007) Designing the myExperiment Virtual Research Environment for the Social Sharing of Workflows, *Proceedings of the Third IEEE International Conference on e-Science and Grid Computing, Bangalore, India, December 2007*, Washington, D.C.: IEEE Computer Society, pp. 603–10.

[28] Taverna Workflow Management System. Available at http://www.taverna.org.uk.

[29] myExperiment – social networking and workflow sharing environment for scientists, Available at http://www.myexperiment.org.

[30] Shibboleth and UK Access Management Federation, http://www.ukfederation.org.uk/

[31] R.O. Sinnott, T. Doherty, C. Higgins, P. Lambert, S. McCafferty, A. Stell, K. J. Turner, J.P. Watt. Supporting Security-oriented, Inter-disciplinary Research: Crossing the Social, Clinical and Geospatial Domains, in *Proceedings of International Conference on e-Social Science*, June 2009, Cologne, Germany.

[32] R.O. Sinnott, D. Chadwick, T. Doherty, D. Martin, A. Stell, G. Stewart, L. Su, J. Watt. Advanced Security for Virtual Organizations: Exploring the Pros and Cons of Centralized vs Decentralized Security Models, *8th IEEE International Symposium on Cluster Computing and the Grid* (CCGrid2008), May 2008, Lyon, France.

[33] R.O. Sinnott, S. Hussain, Architectural Design Patterns for Security-oriented Workflows in the Social Science Domain, *Proceedings of International Conference on e-Social Science*, Cologne, Germany, June 2009.

[34] S. Hussain, R.O. Sinnott (in press) Security-oriented Workflows for the Social Sciences, *4th International Conference on Network and System Security*, Melbourne, Australia, September 2010.

[35] X.D. Wang, M. Jones, J. Jensen, A. Richards, D. Wallom, T. Ma, R. Frank, D. Spence, S. Young, C. Devereux, N. Geddes. (2009). SARoNGS: Shibboleth Access for Resources on the National Grid Service. In *Proc. 5th Int. Conf. on Information Assurance and Security* (IEEE Computer Society, 2009)