

National eInfrastructure for Social Simulation (NeISS)

Community Scoping, Outreach & Impact Report

Meik Poschen, MeRC

Final Version, August 2012 & Minor Revisions, March 2013

With contributions by Rob Procter, Mark Birkin, Alex Voss, Paul Lambert, Andy Hudson-Smith, Andy Turner, Nicolas Malleson, Tobias Schiebeck, Kenny Baird

1. Introduction and Aim

This report is being delivered within the ‘WP4.2 Community Engagement and Capacity Building’ of the JISC funded National eInfrastructure for Social Simulation Project (NeISS). The aim of this work was to scope the community for needs, raise awareness and encourage adoption of new social simulation models and tools developed in NeISS and gather findings, establish contacts and build capacity for future use and further activities. ‘WP4.2.1 Outreach activities’ includes community scoping, requirements gathering and collaboration with some of the WP4.1 Exemplars to this end. ‘WP4.2.2 Training and Capacity Building Resources’ is complementing WP4.2.1 with training materials and activities as well as capacity building via website and Social Simulation Tool development. A third part was looking into sustainability of NeISS under ‘WP4.2.3 Sustainability’ – this work is covered by a separate report¹.

2. Approach

The NeISS project is strongly embedded within a number of different projects and activities, chapter 3 sheds some light on its foundations, community and linkage with a focus on general outreach and impact. Table 1 below shows the potential for NeISS’ outreach areas based on previous work and established communities and contacts through all project partners by mid-project. The concrete work conducted for this report and its underlying work packages is aligned along the five target user categories (1.Research, 2.Teaching, 3.Policy-makers and planners, 4.Public, and 5.Business Organisations) and is discussed in chapter 4.

This outreach and community scoping work started with collaborating with the NeISS exemplars in WP4.1, especially with ‘WP4.1.1 Epidemiology’, ‘WP4.1.3 SimCity for Real’ and ‘WP4.1.4 Ageing and Inequality’. The notion of those three exemplars has been to develop tools within NeISS for a specific application area, driven by the exemplar owners as the domain experts. For our outreach activities this meant to liaise closely and take the opportunity to scope the relevant community, feedback user requirements to the developers (in terms of envisioned/potential use, functionality and benefits of such NeISS tools for their domain/community) and establish an understanding of what the general and specific needs for a social simulation infrastructure will be. The other three exemplars (‘WP4.1.2 Virtual Exhibition Space’, ‘WP4.1.5 Credit crunch’ and ‘WP4.1.6 CityDB’) are public outreach demonstrators developed by CASA and will be covered briefly under general outreach activities (chapter 3).

To cover all of the five outreach areas further activities included to follow up on established contacts (DUG), establish new ones (TfGM, Oldham Council) and make use/create outreach and training

¹ Procter, R., Poschen, M. and Voss, A. (2012). NeISS Sustainability Report.

materials (website, videos, Population Simulation tool, Social Simulation Course) to raise further awareness, gather insights on the usefulness of social simulation infrastructure components, evaluate developed tools and training modules.

Type of User	Size of user community	Why interested in NeISS outputs	Current functionality	Contacts
1. Research a. Geography b. Criminology c. Health sciences d. Social Policy e. Economics f. Transport g. Town Planning h. Engineers i. Research Institutes 2. Teaching a. Higher Education b. Secondary Education c. Professional organisations 3. Policy-makers & planners a. Schools b. Housing c. Roads/ transport d. Economic development e. Work and pensions f. Land-use planning g. Health and social services h. Police i. Central Government j. Non-Government Organisations 4. Public a. Individuals b. Community organisations c. Special Interest Groups 5. Business Organisations a. Retailers b. Financial services c. Property developers d. Consultants e. Utilities f. Insurance companies g. Data providers h. Intermediaries	N=10 ²	Evaluate the impact of policies and strategies Target resources Explore relationships between activity and environment	REPAST MASON ESRI MapInfo MIMAS ESDS EDINA TranSims EpiSim SPSS/SAS/Stata	Met Office, UK Water Industry Research Universities of Manchester, Leeds, Southampton, Stirling, Glasgow, London. Daresbury. Macaulay. Liverpool. Sheffield. Villach. Arizona State, University of Illinois. Argonne. BGS. George Mason. NCG Maynooth. University of Paris. NATSEM (Canberra). DCC, LUDOS Jean Sykes (British Library). AGI. Institution of Civil Engineers, Institution of Mechanical Engineers, Institution of Engineering and Technology.
	N=10 ⁴	Search for evidence of spatial and temporal trends Identify opportunities	SimCity SimModel	DEFRA, DfT, DCLG, DECC, Cabinet Office, Infrastructure UK, Environment Agency, Highways Agency, Transport Scotland, CABE, Association of North East Councils, Hampshire County Council Mayor of London/LCC/ GLA Economics Leeds City Council Social Services/ Chief Regeneration Officer/ Chief Housing Officer Trafford; Telford; Barnsley; Bradford DWP. North-West Regional HA. Safer Leeds. Mersey Police. Yorkshire Forward; Acxiom User Group. Town and Country Planning Association, Local Government Association, Royal Town Planning Institute. Home Office.
	N=10 ³	Early warning systems for problems Performance assessment Justify decisions Scenario-based planning Optimise delivery	LARIA Dr Foster MVA Saturn UrbanSim	SurveyMapper. MapTube. Sainsbury's; Asda-Walmart; Tesco; Planet Online. GMAP; CACI; Experian; OS; Acxiom. Scottish and Southern Energy, E.On, BP, National Grid, Network Rail, United Utilities, Yorkshire Water, Northumbrian Water, Veolia, BT. AEA Technology, Arup, Atkins, Black and Veatch, Halcrow, JBA, Mott MacDonald, MWH, Royal Haskoning, Parsons Brinckerhoff, Swanbarton. Willis. Barn Nuttal, Costain Demographic Decisions Google. PLP Architecture. Volterra Consulting.
	N=10 ⁷	Visualise distributions	Wikipedia Ask Google	
	N=10 ³	School of Geography, University of Leeds Letters of Support (NeISS/TALISMAN/GENeSIS). Infrastructure Transitions Research Consortium GENeSIS/ TALISMAN NeISS Partners	OS CACI Experian Acxiom Verdict	

Table 1: NeISS Outreach Areas (Mark Birkin)

3. NeISS Project Community & Linkage: General Outreach and Impact

Quite a number of projects and activities are connected to NeISS (e.g. starting with MoSeS, GeoVUE, & GENeSIS; CASA activities like SurveyMapper and MapTube; TALISMAN; details see below) and have helped to constitute and conceptualise the project, its evolution and the continuation (post-project) of ideas, methods, outputs and tools. This chapter will give a brief overview on the general outreach and impact created in this context, also including NeISS work not directly covered by this report's underlying work packages. Some activities (methods, models, tools) would not have come into existence without NeISS, at the same time NeISS benefited from the work done in connected projects.

The main drivers in that sense for the emanation of NeISS lay in previous successful work and therefore pre-existing networks and people of research groups at Leeds, London, Stirling, Daresbury and Manchester and a number of projects and nodes of the National Centre for e-Social Science (NCeSS), sponsored by ESRC at the time, namely

- The Modelling and Simulation for e-Social Science (MoSeS, <http://www.ncess.ac.uk/research/geographic/moses/>) project and the Centre for Spatial Analysis and Policy (CSAP, <http://www.geog.leeds.ac.uk/research/csap>), University of Leeds. The foundation for a number of methods and models has been laid here to be further developed, re-used and refined in GENeSIS, NeISS and TALISMAN.
- The Geographic Virtual Urban Environments (GeoVUE, <http://www.ncess.ac.uk/research/geographic/geovue/>) project and the Centre for Advanced Spatial Analysis (CASA, <http://www.bartlett.ucl.ac.uk/casa>), University College London (UCL). The work on tools like the GMAP Creator (and other software:

<http://www.bartlett.ucl.ac.uk/casa/latest/software>), MapTube (<http://www.maptube.org/>) and SurveyMapper (<http://www.surveymapper.com>) started here.

- GeoVUE and MoSeS then became GENeSIS: GENerative E-SocIal Science (<http://www.genesis.ucl.ac.uk/>) in a further phase of ESRC funding, with CASA and CSAP as leading centres of expertise in spatial modelling and simulation collaborating even stronger.
- The Data Management through e-Social Science (DAMES, <http://www.dames.org.uk/>) project and the School of Applied Social Science (<http://www.dass.stir.ac.uk/>), University of Stirling.
- STFC, Daresbury Laboratory (<http://www.daresburysic.co.uk/about-us/campus-partners/daresbury-laboratory.aspx>); provider of the Sakai Research Support Portal (<http://portal.ncess.ac.uk:8080/portal> ; formerly NCeSS portal).
- Manchester eResearch Centre (MeRC, <http://www.merc.ac.uk/?q=node/338>), successor of the National Centre for e-Social Science (NCeSS) Hub, University of Manchester. Also: The NCeSS e-Infrastructure for the Social Sciences (NeSS) project, managed by the NCeSS Hub and funded by the ESRC.

All of the underlying institutes also became NeISS project partners (here the complete list: <http://drupals.humanities.manchester.ac.uk/neiss3/about-neiss>) and most of them are involved in complementing activities relevant to a social simulation eInfrastructure since then; in this context the following list shows a selection of relevant communities and areas connected to pathways for existing and future outreach and impact:

- **FuturICT** (<http://www.futurict.org.uk/> ; <http://www.futurict.eu/>) is a **European initiative** preparing a flagship proposal “to understand and manage complex, global, socially interactive systems” while “integrating ICT, Complexity Science and the Social Sciences”. Making use of methods and tools for forecasting, modelling and social simulation is part of this agenda.
- The **Infrastructure Transitions Research Consortium** (ITRC, <http://www.itrc.org.uk/>) “delivers research, models and decision support tools to enable analysis and planning of a robust national infrastructure system”.
- **Research methods & teaching: TALISMAN** is a node of the National Centre for Research Methods (NCRM), funded by ESRC (<http://www.geotalisman.org/>) and again involving both CSAP and CASA. Between 2011 and 2014 it is continuing the pathway for impact for research and teaching within the NCRM (<http://www.ncrm.ac.uk/>) network on methods for geospatial data analysis and simulation. Further development and deployment of NeISS services is planned through the Talisman training programme.
- **GEOG2080 teaching:** Simulation tools of planning decisions for use with u/g students have been developed at CSAP (class of 110 students; in the future it could be explored how to scale this to a social simulation infrastructure and across institutions).
- **Teaching & cloud computing:** The problem of scaleability is being explored through the ABSS project (cloud computing), which is using NeISS services as social simulation exemplars for a postgraduate community and to encourage uptake of these methods.
- **PhD project on methods:** An ESRC studentship at CSAP (“Geodemographics: Creating a classification at the finest spatial scale”) is using NeISS models as a platform for spatial analysis of outcomes.
- **PhD project on Health:** An ESRC TALISMAN studentship at CSAP is looking into “Modelling the impacts of demographic ageing on the delivery of health care services”.
- **Criminology:** The JISC funded geocrimedata project (<http://geocrimedata.blogspot.co.uk/>) is bridging from academics to practitioners in this research domain by “taking existing geospatial data making it useful for crime analysis”. NeISS models and tools could be further tailored to work with this community.
- **Housing** (housing benefits): CSAP is working with Leeds City Council on scenarios and simulations.
- **Crowd-Sourcing:** An upcoming/new project collaboration between CASA and CSAP includes Twitter work and new modelling approaches based on NeISS.

- **CASA Conference:** CASA regularly (usually yearly) hosts a conference especially aimed at policy makers and the public in general (free to attend, around 200 attendees on a regular basis).

To illustrate how different projects, activities, communities and people can interlink, here a more detailed example from Andy Turner (University of Leeds, <http://www.geog.leeds.ac.uk/people/a.turner/>) and Tom Doherty (National eScience Centre, Glasgow) on their demographic model development and simulation work involving NeISS, GENeSIS and GridPP (as of March 2012):

“The demographic model development work was largely GENeSIS work. In NeISS we have refactored the code so that the programs will run as part of a ‘Genesis Simulator’ that uses GridPP site data storage and compute elements. We have tested that the system works and have produced individual level demographic simulation results for Leeds. These we still regard as ‘test results’ and they were presented with details of the e-Infrastructure at the UK e-Science All Hands Meeting in September 2012. The paper was reworked and submitted for the special issue of the Royal Philosophical Transactions A and we are currently awaiting peer review feedback.

The GENESIS model is designed to input data in a standard XML format and can produce simulation results for any place so long as the data is prepared according to the schemas for the XML. For UK runs, we use Census data inputs and ONS Vital Statistics to generate the Population, Mortality, Fertility and Miscarriage probability data.

We are currently making some changes to enhance the model and outputs and are preparing data to generate simulation results for England from 1981 to 2011. We hope that these results will be of more interest to Demographic Modellers (than the results we have produced for Leeds). As yet the demographic modellers we have tried to engage with have found it curious that we are attempting to model at a daily time step and they seem sceptical that this can be done and that the results will be interesting or useful.

Looking forward, we are planning to submit a proposal to the ESRC Secondary Data Analysis Call and to gear the models to look specifically at health. We aim to reach further to link with epidemiology and health care planning in this next phase. We are still building our pathways to impact, but we believe that there is much potential in this direction.”

Activities directly connected to other NeISS work packages:

- **Transport Scenario:** NeISS developed a use case for transport², which has been used to illustrate the project’s work in a number of user and outreach meetings (e.g. with TfGM).
- **SurveyMapper** (a free real-time geographic survey and polling tool developed at CASA: <http://www.surveymapper.com/>) and as an especially successful example the **CreditCrunch Survey** (Exemplar WP4.1.5: <http://www.surveymapper.com/mapView.aspx?id=2>): The Credit Crunch Survey provided pathways to impact and public outreach examples of research services. Excellent outreach with 22,000 inputs was achieved in providing an early example of crowd sourcing in association with BBC Radio 4. The survey ran for a week providing strong coverage of the research and detailing the real-time mapping output of this work. Other successful surveys included **5Acts for WildLife** (<http://www.surveymapper.com/5Acts> ; January 2011) an Animal Welfare Campaign Twitter Survey and the **BBC Look East Broadband Survey** (<http://www.surveymapper.com/mapView.aspx?id=172> ; January-February 2011) to measure Broadband speeds across the East of England. Furthermore the crowd sourcing of data and model outputs enabled the extension of the Credit Crunch exemplar for other crowd sourcing outputs (<http://www.surveymapping.com/>). A number of CASA activities in this area included liaising with the Mayor of London’s office.
- **CityDB** (exemplar 4.1.6: <http://citydashboard.org/> - for more details also see <http://oliverobrien.co.uk/2012/04/citydashboard/> from its launch April 2012) is an endeavour by CASA providing an online City Dashboard/Database for viewing live data feeds.

² Birkin, M., Procter, R., Allan, R., Bechhofer, S., Buchan, I., Goble, C., Hudson-Smith, A., Lambert, P., De Roure, D., Sinnott, R. (2010). Elements of a computational infrastructure for social simulation. Philosophical Transactions of the Royal Society, Series A, 368(1925), 3797-3812. <http://dx.doi.org/10.1098/rsta.2010.0145>

- The **Virtual Exhibition Space** (CASA; exemplar 4.1.2) is an online space for the display of NeISS output which is still ongoing – with a possibility to move to ‘Lumion’ allowing high quality output of an exhibition space but in movie form. See <http://www.digitalurban.org/2012/02/data-space-agent-based-models-sketchup.html> for the current example.
- Related CASA activities (**Tweet-o-Meter**: <http://www.casa.ucl.ac.uk/tom/>) and general public outreach success: CASA developed social networking tools and analysis approaches leading to excellent impact with coverage of the twitter mapping on CNN, Metro, Daily Mail. The British Library in London commissioned an analogue ‘Tweet-o-Meter’, it hung on a wall in the library for a year to communicate research into social networks.
- **myExperiment** (<http://www.myexperiment.org>), a social curation and community website for researchers “to find, use and share scientific workflows and other Research Objects” set up a social science instance for NeISS (<http://neiss.myexperiment.org/home>).

4. Community Scoping: Outreach Areas & Impact

4.1 Research

Social Simulation Core Domain and Exemplar: ‘Sim City’ for Social Sciences – Population Simulation Tool

Domain Expert Interviews

Eight telephone interviews (average length 33 minutes) have been conducted in March/April 2010 with social simulation domain experts with the aim to determine base requirements for the NeISS infrastructure, functionality and data, to scope the social simulation expert domain and to get insights towards potential benefits of using NeISS.

Setting

The domain experts come from an international community and besides the UK included users from Ireland, Australia and Austria. They have been approached via the Leeds NeISS Social Simulation experts who also belong to this community and collaborate with most of the interviewees on a regular basis or have done so in the past. The Social Simulation sub-disciplines covered by the domain experts are social & economic, health, agriculture, environmental, innovation and transport/travel while they are mostly working with small area or general spatial microsimulation, both static and dynamic. They work nationally and internationally, in part bilaterally and in EU contexts with diverse datasets and are particularly interested in developing and comparing methods, models, outcomes and policies pertaining to social simulation, not only in the academic sector, but also in terms of doing consulting for policy decision makers.

Impact

The following main benefits of a Social Simulation Infrastructure like NeISS have been pointed out:

- Provide web-based access to a general infrastructure and resource for microsimulation (multi-functional; joined-up frontend and backend) providing generic tools and data input/output functionality.
- It is important to have an ‘easy’ usable interface for experts and non-experts from academia and policy decision making/consulting to run microsimulations (using various datasets and algorithms).
- Foster comparison of different methods, models and outputs (also internationally long term).
- Provide links to resources, documentation (wikis) and publications.

At the same time a number of recommendations and potential caveats have also been identified:

- The validation of models, methods and algorithms is essential, things have to work as expected and people have to trust the expected outputs (important for uptake) – “It would make sense to give the user the option of running the model for a small area and test the result.”
- A lack of understanding (of the process and/or functionality) by non-expert users could lead to incorrect results due to too complex functionality

- The issue of time in doing microsimulation: how fast will the portal be able to work (with the underlying infrastructure)?
- To address policy makers can be quite ambitious due to the way they work: usually there are 2-3 hours to present something – and with questions coming up, unless the system can do things in real time, another driver, incentives are needed for adoption by consultants.
- Trying to cover a number of methodologies and datasets will be useful and help adoption.
- Approaches to access right policies, IPR, open access and cost models have to be considered.

Population Simulation Tool

The NeISS Population Simulation tool

(<http://drupals.humanities.manchester.ac.uk/neiss3/content/neiss-population-simulation-guided-simulation>) has been developed based on Leeds' social simulation models lead by MeRC and in collaboration with Leeds (social simulation expertise and provision of computer cluster), Daresbury (provision of Sakai portal and server/database infrastructure), CASA (MapTube) and Stirling (additional variables via Age & Inequality exemplar and related work packages). The tool is a guided simulation within the Sakai portal allowing the user to run through a complete simulation workflow for a selected area of the UK to be chosen via a map and visualise the results through MapTube. This is intended for anyone who is interested in trying out their own simulation, guiding through the process, while providing the opportunity to control many of the variables. The simulations are guided in the sense that there are limited variables and therefore pathways to choose, but the actual computation is run live (if a previous result is not already saved and therefore provided 'canned'); all output files are available for download after each step is completed and the user can go forwards and backwards through the different steps and results, run new selections and go back to completed ones. The Population Simulation includes the following four phases:

- I. **Map Population Reconstruction:** The user can choose one area or a selection of areas on a map based on Census data to create a starting population of individual households. This Population Reconstruction Model (PRM) provides finer statistical detail about the included people than just using Census data sets. Three degrees of quality are available, coarse (based on a set of 100 people which is then extrapolated to create the population), medium (a few thousand people records are used) and rich (the full data set with half a million records is used).
- II. **Dynamic Model Generation:** The Dynamic Simulation Model (DSM) uses the (in the PRM) created set of synthetic households and simulates the processes of birth, death, marriage, reproduction, migration etc. to predict how the regional population will change over time. In this manner it is possible to update an old census or make future predictions about household structure. The user can enter the number of years the population will be projected into the future, e.g. 15 years would estimate the synthetic 2001 population as to how it might look like in 2016.
- III. **BHPS Linkage:** This step allows the user to choose additional variables based on the British Household Panel Survey (BHPS) which will be linked up with the Census data categories, e.g. age, to enhance the results with more information categories like income, if a person smokes etc. These common characteristics will then be further included in the population estimates.
- IV. **Mapper Service:** Finally the generated data is mapped, i.e. visualised using a query to MapTube which gives back a map with semi-transparent overlays, coloured per data characteristic. The user can select the value to be mapped as well as percentages or absolute values. Like all results the mapping also is provided as a downloadable file (CSV format).

Impact

Requirements and insights gathered in the social simulation domain expert user interviews (see previous section) had a significant impact on the prerequisites for its development. The tool was used as a demonstrator across various disciplines and domains as further detailed in the respective sections of this report (i.e. within the 'Age and Inequality' exemplar, in the meeting with Oldham Council, and as a training video). Feedback collected in those cases was also put to good use for further improvement, especially on the help/guidance texts and on usability. It is planned to use the Population Simulation tool as one main means for further future outreach and demonstration activities. At the moment the simulation uses the 2001 Census data, it is intended to also include the recent 2011 Census data when applicable.

Age & Inequality Exemplar

The Age & Inequality exemplar focused on using social simulation for sociological and econometric research agendas studying the circumstances of social inequality and hence developing methods to forward-project occupational and educational qualification structures in this context (social mobility and inequality: ageing population; changing family structures; educational expansion; immigrant influxes; wellbeing) using long running longitudinal survey data resources in UK, e.g. BHPS. An interview with Paul Lambert (October 28, 2009), the Age & Inequality exemplar owner in the NeISS project was conducted to get first insights on the specifics and scope of the domain and its community in relation to NeISS, with the documentation of this session being further discussed and refined in the following months. A NeISS ‘fringe workshop’ session on ‘Simulation Analysis of Ageing and Inequality’ was then held at the ‘Social Stratification Research Seminar’ at University of Stirling (August 31, 2011) to present the NeISS approach and give a demonstration of the NeISS Population Simulation Tool (see section above; at that time the tool was still under development, feedback from the WS session and interviews was incorporated) also used for to show how social simulation can be useful for this specific domain and learn from an initial discussion with the domain experts. The event was further followed up by five telephone interviews (November/December 2011) with workshop participants to elicit more detailed feedback on their view on NeISS and potential benefits and impact of social simulation for their research communities.

Setting

Both economics and sociology are as related disciplines interested in social mobility and inequality, especially towards labour market structures and educational qualifications, but until now not much work has been done to try to forward-project the evolution in those structures. Some of the research questions within this exemplar which have been put forward to the community for discussion were (examples from NeISS ‘fringe workshop’, slides³):

- “How age qualifications links impact trends in social inequality – Mass education; admissions policies; cognitive versus sheepskin effects”
- “How will (high/low qualified) cohort specific immigrant influxes impact upon regional age occupation qualification distributions – Simulation: increase or decrease proportions within birth cohorts/ethnic groups/regions/sectors with certain qualifications”
- “How will fine grained industrial sector transformations impact different age cohorts and subsequent stratification positions (e.g. rise of the ‘cultural industries’) – Simulation: Modify national and/or local industrial distributions and project forward over time”
- “How is long term wealth accumulation influenced by longer life expectancies (e.g. changing inheritance patterns; longer pension dependence) – Simulation: Model and modify income through work and through inheritance as it influences relative social position at a national level”

Three potential audiences in this domain have been initially identified, acknowledging that reaching those will be ambitious: 1) the largest group would be interested in general outputs as the results of social simulation; 2) a relatively small group of quantitative sociologists/economists interested in the general field might be persuaded to try out a simulation style analysis and to replicate or modify the NeISS approach; 3) groups interested in the methodological results and inputs of NeISS (quantitative sociology, social survey research, e.g. “typically CCSR type people”). The five interviewees use or have used classical quantitative statistical models and mostly longitudinal surveys/data for their research, but none used simulation techniques so far.

Impact

- Most workshop participants and interviewees agreed that making the effort to bridge the disciplinary gap between the social simulation community and other domains towards making use of social simulation tools and methodologies – especially those already using quantitative statistical methods – would be a worthwhile endeavour. Events like the conducted NeISS ‘fringe workshop’ can be suitable means to raise awareness in this context. The following points have been emphasised:

³ Lambert, Paul; Poschen, Meik; Warner, Guy (2011): Simulation Analysis of Ageing and Inequality. A ‘Fringe workshop’ (within the NeISS project) to the ‘Social Stratification Research Seminar’, University of Stirling, 31 August 2011. Slides available under http://www.merc.ac.uk/sites/default/files/neiss_session_31aug11.pdf

- The specific terminology around social simulation and the benefits of using such an approach have to be made clear cross-discipline: What is social simulation? How can social simulation help solving research questions and to what end? It should be presented as an additional means to tackle problems, give indications and present trends, another way of describing social phenomena, not as an exclusive or absolute method.
- One potential benefit of the tool lies in it being useful to visualise geo-located data, as in “a presentation of explaining results”, “an easier way to explain”.
- Offering different scenarios in the case of more static tools and also enabling the use of domain specific datasets would be useful – but also will present a challenge tailoring it to every research community.
- Proper documentation and references have to be provided of the methodology as such, the algorithms, results and functionality of the tool(s) involved.
- Validation and transparency of method and results is crucial, a kind of audit trail would be very useful (together with references and documentation) to enable reproducing results.
- Such an approach has to be easy enough to use, meaning the effort and time to be invested and the perceived benefit has to be in balance for the user.
- Additionally personal exchange and collaboration between social simulation and domain experts might in the end be necessary to convince academics from other disciplines, foster uptake, understanding and re-use.

At the same time some caveats have been identified:

- It might be hard (at first especially) to reference a new methodology coming from a different discipline, also depending on the general time constraints to familiarize oneself with new approaches properly and on the concrete research practice and culture – the points made above can help mitigating this barrier and open up communities for social simulation.
- The use of terminology like forecasting, prediction and social simulation is sometimes seen as too absolute or dubious within social science/sociological paradigms (sociology is much more historical as a whole), where the question of how to measure social processes always is under debate. If predictions are made, “there needs to be some sort of measure of both kind of the accuracy and also the kind of consequence”.

Epidemiology Exemplar (Health Research): UK Disability Estimates via MethodBox

Within the outreach work packages MeRC facilitated the development of a NeISS tool in Epidemiology at the University of Manchester, an exemplar activity which came to fruition only in the last months of the project. The NeISS UK Disability Estimates tool (<http://drupals.humanities.manchester.ac.uk/neiss3/content/uk-disability-estimates-methodbox>) is based on MethodBox (<https://www.methodbox.org/>) and prepares a Health Survey for England (HSE) 2000 subset which can be used along with a prepared HSE 2001 data subset and cut down Census data info to calculate disability prevalence for 2001 to 2031 in UK districts. The script uses the R statistical programme and is based on the Small Area Estimation work by Dr Alan Marshall (available from <http://www.ccr.ac.uk/esds/data/>).

Impact

It is intended to use tool and established collaboration for further future outreach in the Health Research domain.

4.2 Teaching

Social Simulation Course

Within the Digital Social Research (DSR; <http://www.digitalsocialresearch.net/>) funded project ‘Developing Sustainability Pathways for Social Simulation Tools and Services’ NeISS was one of the partners involved in developing training materials and running a five day Social Simulation Course module at the North West and White Rose Doctoral Training Centres (DTC) in April/May 2012 (see <https://sites.google.com/site/socialsimulationcourse/home>).

Setting

The course was primarily aimed at Social Science postgraduates and research staff and focused on encouraging participants to think through the possibilities for applying social simulation in the context of their own research. The course introduced the methods and techniques of social simulation, including microsimulation and agent-based modelling and discussed a number of detailed case studies ranging across the social sciences.

Impact

25 participants mainly from Manchester and Leeds attended. The course was received well and rated very positively by the participants with 14 course evaluation forms being returned (for a summary of the evaluation form results see

<https://sites.google.com/site/socialsimulationcourse/home/ISS%20Evaluation%20RESULTS.pdf?attredirects=0&d=1>). Some highlights of the evaluation results are:

- The course content was considered as being ‘quite’ to ‘very interesting’ and the course material as high quality, with the interaction (opportunity to ask questions) between participants and course tutors/lecturers rated as excellent.
- 12 out of 14 attendees expect to use social simulation in their future work.
- A third of the participants found the degree of difficulty challenging (one found it outright difficult due to perceived “high programming demands”), but the speed of presenting the concepts was considered to be about right with a fair to excellent clarity of explanation (only one rating was below fair here).
- The format of the course was seen as beneficial; nevertheless it became clear that one half of the participants were more adept and interested in programming than the other – a need which should be catered for better in the future.

Course design and findings are currently assessed further within the ‘Developing Sustainability Pathways for Social Simulation Tools and Services’ project towards running the course for more audiences (other DTCs, NCRM & Talisman context, further market research), exploring the distant learning angle, and for offering a portfolio of various formats for such a course in the future and about the implications.

4.3 Policy Decision Makers & Planners

Transport for Greater Manchester (TfGM)

NeISS arranged a first meeting with TfGM (July 8, 2011) with the premise to learn more about the kinds of forecasting work that TfGM undertakes, how this is currently done and what kinds of challenges it faces for the future in order to identify and explore possible ways in which the NeISS project might be relevant to TfGM. A follow-up meeting was then held in autumn (October 24, 2011) to further discuss mutual interests and benefits for collaboration. Based on those activities NeISS contributed a letter of support to TfGM’s successful bid to the DfT’s Local Sustainable Transport Fund (LSTF).

Setting

TfGM are interested in fostering collaboration with academia, they would see it as beneficial to explore opportunities within their remit with a focus on Greater Manchester (GM) as a region. The GM strategy strongly envisions a synergetic approach for the region (various commissions on transport, health, economy etc. play an important role) which means TfGM increasingly has to collaborate and exchange data with other GM bodies. The difficulty with making full use of such a multi-agency approach lies in the current lack of (quick) solutions to really access, use and aggregate different data (sources) from different agencies. In this context one of the main themes at the moment, ‘low carbon economy’, poses the question of how to represent the impact of background changes in our economy: “how can we do something now which tests our assumptions”. TfGM is part of a multi-agency interdisciplinary group which tries to address these questions from a broader perspective in order to build a common understanding of the problems (understand peoples view, their assessment of what they can do themselves) and reach policy interventions.

TfGM are using a number of conventional, more static models for transport looking at interventions and developing business cases, but would like to improve decisions making based on more dynamic

modelling and simulation with the caveat that they find it problematic to reduce complexity and contingency to that end. One example would be looking at developing a climate change profile: based on available data, how to set behavioural parameters sufficiently and put those into a model? “How do we represent complexity in real life?” Lifestyles of people have to be in the centre of this process to define those behavioural patterns, which means dealing with something very complex vs. working on the base of very static resources (and the source data also plays a role in determining how static those resources are). TfGM is also exploring how to get into a rolling programme of getting and using data more incrementally, in shorter cycles. One activity here is about creating and assessing long term trend profiles of the average user in regard to specific areas, e.g. putting together a travel diary over a year’s time. This involves monitoring data of how people really use transport and this kind of data it is very hard to get. Also the demand side (customers) is seen as important, i.e. how to get information/data on what demands are really out there, how are they configured and might be influenced by planning and measures. The general difficulty lies in building a picture on the micro-level and then putting it into the larger picture to help with decisions. Also: “The policy agenda of choice is the massive added complexity.”

Impact

NeISS contributed a letter of support to TfGM’s bid to the DfT’s Local Sustainable Transport Fund (LSTF), which secured £32.5 million funding end of June 2012 (http://www.tfgm.com/tfgm_news.cfm?news_id=9009337?submenuheader=4) focusing “around the three core themes of active travel, smarter travel information and promotion, and network efficiency” (<http://www.tfgm.com/ltp3/LSTF.cfm> – includes supported ‘revised full business case’ of bid, submitted 20th December 2011). This includes topics like bicycle commuting and ticketing information systems. LSTF consists of a portfolio of initiatives and activities with a wide range of measures. Within TfGM’s remit of exploring how to gather and use data more incrementally and in shorter cycles to improve on the ‘static data’ issues one focus of the LSTF initiatives lies in collecting (more) real life data, looking at the ‘softer issues’ and implications. Manchester’s Media City would be an example of bid initiatives concerning improving the Metro Link route, supporting local access routes, and marketing infrastructures for mass markets. For collaboration with NeISS: “[...] the area of most interest to TfGM at the moment would be exploring how your techniques could help analyse behavioural responses to sustainable travel initiatives. [...] one item for discussion would be how we could jointly develop a research proposition for the bid, which could sit alongside the programme over the next 3 years or so.” It is important for TfGM to find out to what extent the respective problem is clear, how can the understanding of markets, customers, people be fostered from the already existing (but somewhat static) forward projections of historical data to the understanding of primary drivers (behaviour).

Further meetings with TfGM to discuss collaboration on the successful LSTF bid are currently being prepared.

A number of areas have been identified for TfGM to benefit from NeISS in further collaboration:

- Employed forecasting/planning techniques and methods in existing TfGM projects could benefit from opening up the conventional approach with its slightly more static models. Agent-based modelling should be a useful approach for addressing behavioural patterns; at TfGM Agent-based/Activity modelling is not really applied at the moment, because of its complexity and limited in-house resources: “all is more in the here and now, day-to-day” business.
- NeISS can offer taking already collected/available data again and feed it into a given simulation, therefore exploring a range of implications, a clustering, “tipping points”, hints towards shifts in behaviour.
- Furthermore regarding models and complexity, TfGM would benefit from insights on how to educate decision makers (outputs from the 3S project on developing a social simulation pilot module for training might be especially useful in this context as a baseline model) on how to configure models and read findings from the simulation, one example would be the congestion charge – currently the TfGM assumption mostly is to run the model a few times, only to get the average, but calibrating the model more sufficiently would lead to better outputs. NeISS might provide help on: How many dimensions could be put into a model, how many worlds do you need to be planning for, what are the futures to be looking for (due to complexity decisions are actually often made on a hunch, not with strategy), i.e. in the end, how to help TfGM in the decision process towards which future to go for, how to make good informed decisions based on a feasible number of choices? For

transport: what are the determining factors for decisions, besides having fresh and continually updated data available?

- At the base of TfGM policy questions they are looking for “blue sky thinking” provided by NeISS to better understand the complexity of policy issues/questions to make the decision process more successful?
- Generally it would be useful to a) conduct sample case studies, b) do something with students, c) test assumptions and methods with TfGM then making those applicable; d) developing apps for timetables etc. are of interest. Caveat: at the moment there is no formal way to decide on such endeavours within TfGM).
- Potential arrangements for collaboration have been discussed: a) embedding a research under-/post-graduate with TfGM, e.g. to take care of collected data, use specific simulation methods in a field/project; b) vice versa, there is an initiative for ‘real world’ people to work at University; c) good results have been achieved in the past with PhD student working on a housing market segregation model in the East Leeds Area, and d) similarly looking at crime.
- Potentially, maybe out of remit for NeISS itself: There is a need for structured evaluation, not only of initiatives/projects, but also for the whole TfGM programme. At the moment there is no distinct strategy how to approach evaluation, assess impact and revenue, this has not been done at TfGM (formally) in the past. TfGM are open to approaches, any collaboration and expertise would be welcome.

Oldham Council (Geographic Information Manager, Corporate Policy and Research)

NeISS met the interviewee on October 5, 2011 after they had expressed interest in the project’s work. At Oldham Council the interviewee has been working on the local information system for seven years and is concerned with tasks and functionality similar to a ‘data observatory’, mapping statistics down to the small area level and providing the figures to people to use. Integrated into this is a facilities model with all the local schools, community centres/facilities, children’s centres, drug and alcohol services, so this can be cross-referenced with the general statistics. They provide a few hundred services and are “trying to do things in neighbourhood clusters as well”. The interviewee’s research team consists of eight members, in the same unit there is also a semi-detached public health team with four more people; additionally there is a separate community safety unit dealing with crime. They are paying someone to do projections as well.

Setting

What NeISS offers “does sound like you’ve got a bit more” in terms of depth, exactitude, modelling then the interviewee is used to. This would especially be beneficial to link up certain data sets and services for modelling and coming up with something more precise, integrating bus routes, populations (ethnicity, migration models) and other data. The data quality is quite important (e.g. data on crime in Greater Manchester is often not very precise due to the way it is reported on paper sheets not as properly as it could for time and location etc.) as is the access to the data. All authorities are using LLPGs – the Local Land and Property Gazetteer, which includes all addresses – but some still have this on paper and not in a database (they have it electronically). Another challenge identified lies in how to use and integrate the multitude of available web-services. Having projections one to two years in advance would be very useful to be able to react to developments quickly (i.e. for migration), but at the same time they need to look at populations with models in a range of 20-25 years (overall “it will vary again according to what particular needs come up at any one time”).

The NeISS transport scenario (see under chapter 3.) is of interest due to the extension of the MetroLink in Greater Manchester at the moment: they try to come to good estimates of usage and travel distances, benefits for businesses, jobs and there like; as far as the interviewee knows there are no real simulations running on all this, at least not at Oldham Council. The paper on transport (REF) has been sent to the interviewee.

The NeISS Population Simulation tool (see under 4.1) was demoed at the meeting and also tried out afterwards by the interviewee.

Impact

- Having a transport scenario simulating around impacts of the extension of the MetroLink in Greater Manchester would be a very worthwhile endeavour to collaborate on in order to “challenge the

models or challenge some of the economic assumptions that are being used across the whole economic spending pattern”. At the moment the NeISS transport scenario is on ward level, so it would have to be broken down further to be useful.

- The interviewee could see potential of the NeISS Population Simulation tool: “Being able to do other data sources would be very, very powerful.” Oldham Council, as other Local Authorities “invest quite a lot every year in surveying the same set of variables (social surveys), so being able to project those ahead reliably” would be very useful. Oldham Council are very interested in topics such as community cohesion (“big issue”), “how people feel about their local areas and crime”, social and digital exclusion, health, mental health and self-reported health as well. They run a sample of a couple of thousand a year on those issues spread across Oldham and try to repeat some of those yearly or each question recurring every three or four years.
- On the same note the NeISS Population Simulation tool is seen as useful as is as a means to compare projections of different areas on various topics like e.g. smoking. The interviewee tested the tool shortly after the meeting and gave valuable feedback for improvements, especially on clarifying its processes, given values and outputs.
- Potential costs for service in the future: the interviewee finds a potential, moderate cost for usage (probably depending on the kind of user, i.e. commercial, academic, and in order to sustain the service) reasonable, although local governments are always short on funds, especially at the moment. Showing benefits and in the end cost savings via such a service can be a way to go.
- The Local Authorities Research Council Initiative (LARCI, <http://www.rcuk.ac.uk/research/xrcprogrammes/OtherProgs/larci/Pages/home.aspx>) and the Local Authorities Research & Intelligence Association (LARIA, <http://laria.gov.uk/>) are communities of relevance for local authorities and likely of interest for NeISS outreach. LARIA is organising an annual conference, with the last event having taken place in March 2012 (The LARIA Annual Conference 2012, Research for Resilience) at Aston University.
- It was tentatively agreed that it would be useful to set up another meeting with the whole team at Oldham Council when the NeISS Population Simulation tool is completed. In this context a more focussed discussion on how to develop/adapt models/simulations beneficial for Oldham Council in particular (e.g. for transport planning/forecasting; potentially crime) is being considered. Integrating surveys and data from surveys used at Oldham would be another way of making use of NeISS. Such could then be used as a proof of concept example to help us understand how to create models and deliver a service that will be useful for other Councils and Local Authorities.

4.4 Public

NeISS Website

The NeISS website under <http://www.neiss.org.uk> serves as the central information hub of the project linking all tangible project outputs. The “Tools” section gives access to all available NeISS tools and associated websites and resources.

Training Videos

For outreach and guidance purposes we have been producing a number of training videos (in close collaboration with the domain experts) for a) Survey Mapper, b) myExperiment, and c) the Population Simulation tool. The videos are envisioned for use as learning resources and for future capacity building activities and are available on the NeISS website (<http://drupals.humanities.manchester.ac.uk/neiss3/content/resources>).

4.5 Business Organisations

Demographics User Group (DUG)

The DUG (see <http://www.demographicusergroup.co.uk/> , also for a membership list covering a wide range of fifteen large and prominent businesses) was set up by our interviewee 1998 to give its member organisations a forum to discuss experiences and good practice on making use of demographic information “to analyse customers, identify markets and avoid risks. [...] to target their resources to achieve maximum effect.” An interview took place on July 13, 2011 with the aim to get an overview of

DUG, to identify business problems that its companies are wrestling with, and to come to ideas on how social simulation and NeISS might be of use in this context, how business models might look like. This was followed up by a presentation focusing on Retail Planning, Demographics and CMA (customer marketing areas), their scope for modelling and potential value for predicting customer behaviour at the quarterly DUG members meeting (September 23, 2011), which led to a lively discussion on the possibilities of developing modelling in new areas.

Setting

Almost all companies have teams involved in market analysis, customer analysis and customer insight, which involves holding huge internal customer file information databases. They are also interested in external, government datasets like the Census, on data.gov and movement surveys, and in geo-located data (mapping) in general. Commercial suppliers of data and know-how are also extensively used. Three main areas can be identified which are common for most.

1) Their in-depth interest and forecasting is quite mature for the whole geography of customers (especially related to their stores and branches) and the competition to come to conclusions on: “Where to locate stores? What should be put in those stores? What are the catchment areas? Who are the competitors? How do we access local markets?”

2) Customer database analysis is used to try and identify characteristics of individual customers, based on their past behaviour in order to predict their current and future needs. This ranges from financial services to the various ‘club cards’ of large retailers to figure out what the customer might be interested in. The notion here has changed to the ‘single-customer view’ in the sense that the history of past purchases or financial services is now linked up under the individual customer’s profile. Address matching is of importance here to allocate transactions of the same person to the same profile.

3) Organising market research surveys is often out-sourced to big market research companies. The main discussion at DUG meetings in this context involves looking at branches, locations, catchments and then cross-referencing the data with the customer files. There is also an increased interest in data from the internet, although this is “rather flakier” quality-wise.

Forecasting and prediction are used in a specific way in this commercial domain as trends are looked at and decisions are made for a quite short-lived cycle; in practice this means analysing only the next day, the next week or month or at maximum three to six months into the future (predicting sales, turnaround, customer types). The knowledge of housing developments (and therefore new customers in an area) also is a factor here, although this seems more down to gathering intelligence in regard to planning authorities and building progress. In the case of companies with fewer but bigger stores additional forecasting in a longer-term sense can be applied to determine if e.g. a specific region will be healthy enough for such a store over the next 10 years. The London 2012 Olympics are one of the exceptional cases (or one-off projects) in which a wider consideration of slightly longer term population/demographic changes is needed by most of the retail sector.

A recurring theme in retail which is becoming more and more important is being aware of population movement 24/7 (i.e. mobility of the population, residential vs. workplace population, people moving around cities, transport). One important example would be the discussion of the catchment area around Edgware Road in London. Over the course of a day a lot of different behaviour by different groups can be identified, for example starting with residents getting active very early, followed by commuters coming in and using ATMs and going to the coffee shops, then disappearing into their workplaces and coming out again at lunchtime to buy things and so on. The interesting point is modelling the ebb and flow of the activities and being able to measure the impact. Mobile phones have become more and more important in this context to acquire data, find hotspots, as is tracking and mapping the bike rides within the London Cycle Hire Schemes. Map mash-ups and in that context areas of classifications are of continued interest. Here new methods and approaches, ‘plausible models or indicators’ are welcome and experience is not that deep as it is with the ‘store forecasting’ mentioned previously under 1) – companies here would look for getting ‘pretty good data’ very quickly in contrast to having ‘perfect data’ after a few months time.

A particular example comes from mobile phone companies, who accumulate a huge amount of data about their customers’ behaviour and some time ago started thinking about a classification of mobile phone users also taking into account their social networks and the probabilities leaving their current provider. Those companies at the moment are busy enough working with the amounts of data they are creating themselves.

Overall it is fair to say that all companies collect rich and important data sets on their own while also looking into external data for their purposes continuously. They are very mature in certain areas of

analysis, but find themselves in a process of continuous exploration in regard to linking up data sets, making use of new methods and in the end understanding how they can best make sense of the vast amounts of data available and benefit from them.

Impact

- Modelling over short time periods is generally needed, longer-term forecasting could be beneficial in cases (specifically for one large retailer), but an offer/consideration of input/discussion should start with the previous. Also, along the notion of ‘predictive analytics’ in the financial services world, a lot is about the assessment of probabilities, where time is not a primary factor.
- New methods and approaches, ‘plausible models or indicators’ for population movement 24/7 in the widest sense (i.e. mobility of the population, residential vs. workplace population, people moving around cities) would be of interest.
- Transport is an important area of interest, like traffic speed, road works, and the general access to areas. NeISS has produced work on transport simulation which might be adaptable.
- Another issue lies in finding the right data sets out of the huge amount of data for example on the data.gov website, but also in regard to the potential that lies within companies’ own big data warehouses. This problem has two facets, a) to be able to find a subset of data sets of interest to be able to realistically work on those, and b) to find better means of access and representation of the data to ‘burrow’ a way in and make sense of it, i.e. with the help of visualisation like Google Maps. It would be interesting to see how/if simulation might be able to draw on some of the data.gov data sets and make them more operational.
- In terms of business models companies do and would buy in services for particular projects or activities they have or which they think might help them. For NeISS this would mean they would have to see a suitable idea, something that captures their imagination to be able to see how one of their needs can be addressed successfully for a potential next step (investing a bit of time and/or money – “they would be very impatient if they didn’t see a hint of benefits pretty early”).
- Despite most of the large companies having their own computer centres in general one of the distinguishing features about NeISS is that it offers the capability of doing large-scale and complex modelling and forecasting which in our current thinking most clients would not be able to provide themselves.
- The presentation at the DUG meeting provided an overview of urban systems and the scope for modelling, and some retail examples from GMAP. This led to a discussion of the possibilities of developing modelling in new areas, with mention of: temporary phenomena, such as changes in petrol prices, and their consequences for store catchments; recession changing consumers’ behaviour (e.g. not going out to buy sandwiches - and hence other products - at lunchtime); simple ‘forecasts’ are no longer enough, flagging the importance of econometrics & time series; impacts, ranging from local changes in store opening hours, to opening a 3rd runway at Heathrow; the relationship with behavioural economics; and consumers’ mindsets “people don’t know why”.

5. Summary & Outlook

The presented work on community scoping, outreach and impact of the NeISS project, its partners and connected communities and related work and the developed/developing eInfrastructure shows the complexity of such an endeavour. Raising awareness, scoping different communities and domains and gathering useful requirements was often a long and winding process, but resulted in a rich picture of how NeISS can be beneficial to users (in Research esp. for Social Simulation practitioners and as a course module, and in the Policy Decision Makers & Planners domain), is seen in a critical light (e.g. by disciplines with a historically different paradigm like Sociology) and has already proven to be successful in the public domain (CASA tools and surveys). NeISS Website and Population Simulation Tool as well as the collaboration with the exemplars have been instrumental in evoking tangible interest and collaboration within the work of this deliverable and for the project.

A number of contacts could be established and new collaborations be started – which opens up routes for (potential) further work post-project (further development of the population simulation tool with Oldham Council; collaboration after successful NeISS supported TfGM bid to the LSTF; using and tailoring the social simulation course for more audiences, contexts and locations). Other activities e.g.

with businesses (DUG network) did not result in a substantial collaboration but evoked interest from all participants which would have to be harvested and explored further.

Future community and stakeholder workshops can be a useful means to follow up and further foster established relationships, continue to bring different communities together, spread the word of the usefulness of social simulation and to try to gather more requirements to improve and develop social simulation tools and evaluate existing infrastructures, especially the NeISS Population Simulation Tool. A number of developments are already supported and funded beyond NeISS, as illustrated in chapter 3 (e.g. TALISMAN, CASA activities, CSAP models, ..).

Overall the gathered community intelligence and obtained insights into users' needs, potential benefits and impact depicted in this report will be helpful in following up with new activities in the context of social simulation.