NeISS: National e-Infrastructure for Social Simulation Service Usage Model (SUM).

Rob Allan
Computational Science and Engineering Department,
Daresbury Laboratory, Daresbury, Warrington WA4 4AD

Contact e-Mail: r.j.allan@dl.ac.uk

Version 1.0

Abstract

This document illustrates the service usage in the NeISS project, National e-Infrastructure for Social Simulation. It shows the overall Service Usage Model plus specific use cases from the project exemplars which are under development.

Contents

1 Introduction

This report illustrates the service usage in the NeISS project, National e-Infrastructure for Social Simulation, which is part of the JISC INF11 Programme. It shows the overall Service Usage Model plus specific use cases from the project exemplars which are under development at the time of writing.

2 Service Usage Model for Social Simulation

This section follows the format of the e-Framework SUM template, see `http://www.e-framework.org/Default.aspx?tabid=608`.

2.1 e-Framework Service Usage Model Name

- Name: Workflow for National e-Infrastructure for Social Simulation
- Alternative Names: NeISS Workflow SUM

2.2 Version

- 1.0

2.3 Version History

Version	Date	Author	Description	Organization/ Project
0.1	25/8/2009	R.J. Allan	Initial Draft	NeISS
0.2	22/6/2010	R.J. Allan	2nd Draft	NeISS
1.0	12/10/2010	R.J. Allan	Final	NeISS

2.4 Rationale

This Service Usage Model describes the workflow architecture, services and applications which are the focus of the JISC funded National e-Infrastructure for Social Simulations project (1/4/2009-31/3/2012). In this project, a user portal plus workflow engine will orchestrate distributed services for security, data access and management, Grid computing and visualisation. The portal interface will also be used to engage a community of collaborating users including policy makers, planners and the general public. Social networking technology (open standards) will be used to manage users and to connect to other repositories such as myExperiment. Shibboleth will be used for primary security (authentication) in the portal via the UK Access Management Federation (as set up by JISC). Sakai is used for the portal framework because of the long experience of our developers (with former JISC funding) contributing to the project and its rich variety of collaboration tools. NeISS surveyed the

leading workflow tools for research and determined that Kepler in the USA and Taverna in Europe were the main contenders with equivalent functionality, see project deliverable D3.3.2. Taverna is therefore used as the workflow engine.

The services described here use transport modelling as an example. They could be applied to other types of social simulation with minor changes. NeISS is intended as a generic infrastructure.

2.5 Classification

To be provided by the submitter:	
SUM Type	[X] Domain [] CORE (a commonly recurring SUM; designation requires e-Framework Integrity Group approval)
Domain(s)	[] Learning and Teaching [X] Research [] Libraries [] Administration [] IT Services [] Common
Maturity	[X] Immature [] Mature
Purpose(s)	[] Exemplar [X] Application [] Modelling [] Toolkit
XOR (exclusive "or")	[X] Service Genres [X] Service Expressions
Development Status	[] Proposed [] Developmental [X] Prototype [] Production
Deployment Scale	[X] Isolated [] Ubiquitous
State Behaviour	[X] Stateful [] Stateless
Transactional Behaviour	[] Transactional and ACID [] Transactional but Non ACID [X] Non-Transactional
Batch Behaviour(s)	[X] Individual [] Batch
Time Constraint Behaviour	[] Hard Real Time [] Soft Real Time [X] None
Service End Point	[X] Provider [] Requestor [] Transcoder (both requests and provides)
Authentication/ Authorisation Dependency	[X] Auth-Dependent [] Auth-Independent
Protocol Binding(s) (only applies to service expression based SUMs)	[X] Web Service [] SOAP [] REST [] HTTP [X] Other
To be determined by the e-Framework:	
Status	[] Approved [] Placeholder [] Unapproved [] Superseded [] Withdrawn
Confidence Level	[] High [] Medium [] Low

2.6 Notation

Services are being drawn from other projects and frameworks as follows. Many of these will have been separately documented on the JISC eReSS Wiki, see `https://www.confluence.hull.ac.uk` or NCeSS Web site `http://www.ncess.ac.uk`.

DAMES: Data Management through e-Social Science `http://www.dames.org.uk`. DAMES is an ESRC research Node of NCeSS. The Node commenced 1/2/2008 and is based at the Universities

of Stirling and Glasgow. DAMES covers case studies, provision and support of data management activities undertaken by social scientists. This includes tasks associated with preparing and enhancing data for analysis, such as re-coding and constructing variables, linking datasets, cleaning data, and data preparation or manipulation. Specialist topics covered in DAMES are data on measures of ethnicity and immigration (GEMDE service), data on occupations (GEODE service), data on educational qualifications (GEEDE service), data on social care, and data and research on social patterns in mental health.

DataMINX: is a collaboration between OMII-UK and NCRIS Australia to develop services for transfer of large datasets. It builds on the Apache Commons VFS software plus the HPC File Staging Profile and OGSA Data Management Interface standards from the Open Grid Forum to provide an open interface to such a service. See `http://www.dataminx.org/` and `http://www.omii.ac.uk/wiki`.

GENESIS: Generative e-Social Science for Socio-Spatial Simulation (GENESIS) is a second phase node of NCeSS funded by the ESRC. The project started on 1/10/2008 and is due to complete on 30/9/2011 `http://www.genesis.ucl.ac.uk/`. The Centre for Advanced Spatial Analysis (CASA) at University College London, and the Centre for Spatial Analysis and Policy (CSAP) at the University of Leeds are collaborating to develop science for the simulation of cities and regions. These two groups have a long history of developing urban and regional models beginning with spatial interaction and allocation models and moving these to aggregate dynamics of urban structure and more recently embracing new approaches at more dis-aggregate levels which build on micro-simulation and agent based modelling. The project will develop a series of demonstrators which will be implemented using e-infrastructures which are emerging from experiments in Grid computing and Web 2.0. The interfaces will be highly visual, building on CASA's GeoVUE and on CSAP's MoSeS projects.

GeoVUE: GeoVUE was an NCeSS Node to develop new applications of Geographical Information Systems (GIS) and related technologies which are focused on 2D and 3D mapping and Computer Automated Design (CAD). The express aim was to develop new kinds of virtual urban environments (VUEs) through which users can further their understanding of cities and engage in other forms of participation. GeoVUE is developing three demonstrators which draw on the applications in related projects. (1) To link publicly available data sets such as those from the National Population Census to non-proprietary mapping and developing some rudimentary exploratory spatial data analysis using geo-demographic data in Google Maps. The mapping hacks that have been developed enable users to take data from an appropriate mapping file and convert it to a form where it can be immediately visualised in a Web page through an interface called MapTube. (2) Focus on linking 2D and 3D map data to real time pollution data. Link this and data which is routinely monitored by the Air Quality Archive to real time GIS. The interfaces will make use of free technologies such as Google Maps and Google Earth. (3) Enable users to fashion their own virtual environment using on-line tools. See `http://www.casa.ucl.ac.uk/projects/projectDetail.asp?ID=57`

G-R-Toolkit: for wrapping applications as Web services, managing data and running jobs on Condor pools or Globus Grids, see `http://tyne.dl.ac.uk/twiki/bin/view/GridAndHPC/GRToolkit`. G-R-Toolkit addresses the security requirements of underlying distributed services. G-R-Toolkit started with the JISC funded GROWL VRE and was further developed in the CQeSS node of NCeSS, Collaboratory for Quantitative e-Social Science, which also developed portlets for use in the Sakai framework.

GridSite: Access control for Web services developed in the GridPP project. Is uses a set of extensions to the Apache web server and a toolkit for Grid credentials, GACL access control lists and HTTP(S) protocol operations. See `http://http://www.gridsite.org/`.

MoSeS: Modelling and Simulation for e-Social Science (MoSeS) was an NCeSS Node funded by the ESRC `http://www.comp.leeds.ac.uk/moses/`. This project completed on 19/12/2008 and work has transitioned to the GENESIS and NeISS projects.

myExperiment: myExperiment is a collaborative environment where scientists can safely publish information. Workflows, other digital objects and bundles (called "packs") can be shared, sorted and searched. Unlike Facebook or MySpace, myExperiment is designed to meet the requirements of the researcher and makes it easy to contribute to a pool of scientific methods, build communities and form relationships based on the use of "research objects". myExperiment is currrently the largest public repository of scientific workflows `http://www.myexperiment.org`.

NeISS: National e-Infrastructure for Social Science `http://www.neiss.org.uk` is the subject of this SUM description.

NGS: National Grid Service `http://www.ngs.ac.uk` provides a set of resources for computational jobs and data manipulation. They also provide services for user and VO management and the UK Certification Authority and MyProxy services. Both NW-GRID and White Rose Grid contribute to the NGS.

NGS Portal: provides a set of portlets for data and job management using JSDL and JSR-168, see `http://portal.ngs.ac.uk`

SARoNGS: Shibboleth Access to Resources on the National Grid Service was a JISC funded project which ran from 1/1/2008-1/1/2009, see `http://www.jisc.ac.uk/whatwedo/programmes/einfrastructure/sarongs.aspx`. The project delivered a standard platform for integrating external resource providers into the NGS using Shibboleth authentication mechanisms. This is in the form of a Credential Translation Service (CTS), developed to work with standard Virtual Organisation Management Services (VOMS) and provisioned with Web service interfaces so that external resource providers can interface with the NGS with minimal development effort. At the Shibboleth authentication level the service delivered in this project interfaces with the UK Access Management Federation and thus provides transparent and seamless access to Grid resources for (academic) users who already have a Shibboleth identity. Open standards used are X.509, SAML and XACML.

Sakai: Foundation provides the second largest open source portal framework used for teaching and learning and also collaboration in research. UK developers have contributed to Sakai in several JISC funded Virtual Research Environment (VRE) projects. Sakai currently supports JSR-168 and Web services in addition to its own internal services and tool framework.

Shibboleth: will be used via the UK Access Management Federation (UK Federation) `http://www.ukfederation.org.uk/`. The Web site provides a number of case studies `http://www.ukfederation.org.uk/content/Documents/CaseStudies`.

Taverna: Taverna is an open source tool for designing and executing workflows. Taverna was developed by the myGrid team and funded through OMII-UK. `http://www.taverna.org.uk/`.

2.7 Description

The National e-Infrastructure for Social Simulation project, NeISS, seeks to provide a production quality e-Infrastructure for social simulation capable of being exploited by a wide range of users which supports the entire simulation research lifecycle and promotes shared understanding of simulation methods, models and results among social scientists. Its specific aims are as follows.

- respond to a growing demand for powerful simulation tools by social scientists, public and private sector policy makers by productionising tools and services developed within NCeSS, the ESRC funded National Centre for e-Social Science, and other relevant e-Infrastructure investments so that they are robust and usable by different user communities;
- develop social simulation exemplars to demonstrate and evaluate use of e-Infrastructure in innovative research within domains of strategic importance. Exemplars will be built around use cases chosen to match needs of different users: 1) inexperienced users accessing services integrated within domain specific portlets; 2) experienced users accessing capabilities at the level of an individual service or workflow; 3) expert users with the capability to introduce their own services and build new workflows through combinations of new or existing services;
- integrate tools, institutional (e.g., repositories, local compute resources) and national e-Infrastructure (e.g., NGS, Edina, Mimas, ESDS, JSTOR) to support the key steps in the simulation research and scholarly communications lifecycle: resource discovery (models and data), data access and management, model building, validation and execution, visualisation of results, curation and publishing of models and results for sharing, discovery and re-use;
- establish standards and frameworks for e-Infrastructure to provide foundations for deployment in discipline areas beyond social simulation;
- work with key stakeholders to: 1) raise awareness of social simulation among research groups, public and private sector users; 2) build capacity and skills through training programmes to enable users to exploit social simulation effectively; 3) provide new services to facilitate public engagement in research and making policy. A steering committee will advise on these activities with membership will be drawn from research (e.g., ESRC National Centre for Research Methods), institutional (e.g., UKRDS) and infrastructure stakeholders (e.g., NGS, Mimas, ESDS, Edina, JSTOR, STFC, JISC). It will also assist in investigation of sustainability issues and devising of business models to match needs of different user communities and stakeholders.

2.8 Business Process Modelling

The MoSeS SUM identified the following business processes (requirements) which are the subject of further development in NeISS. The following description has been updated to reflect the new architecture.

1. Authentication and Authorisation – the access control mechanisms need to facilitate login for academic researchers and potentially commercial partners to set up and run scenarios. Access to certain resources (computers, networks, data sources) will need to be monitored and managed.

2. Social Networking – Some parts of the system will be visible to members of the public or policy makers. They may want to test scenarios, but not have to log on. This requires some form of identity tracking, maybe via a social networking system. Other researchers may want to publish workflows associated with their hypotheses and scenarios for others to re-use.

3. Prepare input data (e.g. upload existing data into the infrastructure); enhance data through additional analysis; access existing derived secondary data (e.g. based on the Census, BHPS, etc., subject to agreement with data provider); collect new data using SurveyMapper to constitute or enhance existing data.

4. Set up and initialisation – there is a first run which comprises a re-constructed population for 2001 at an individual and household level and 30 snapshots of a dynamical model output, one for each year up to 2031 also at an individual and household level. Outputs can be stored for further use in scenario testing. Parameters may need to be added at various stages of the workflow.

5. Generate, adapt and re-use demographic data – new re-constructed populations for 2001 can be generated by specifying a different random seed for the optimisation, or by modifying the optimisation so that it uses a different optimisation function or set of control constraints. If the control constraints are kept the same, or relaxed then a previous result can be used as an initial solution in the optimisation. New dynamic simulations and outputs can be generated by modifying random seeds, or modifying probabilities used in the simulation, e.g. age specific death rates.

6. Aggregation and enhancement of data – linking of further data resources subsequent to data simulation (using same range of potential data sources as (3)).

7. Save population models with metadata and provenance – for re-use and to speed up repeating workflows.

8. Visualisations of the demographic data – generate visualisations of the demographic data at particular times and reveal changes over time. For many applications it is good to be able to show trends and extreme results from a number of different runs of the dynamic simulation.

9. Mapping – visualisation of results is in part achieved through interactions with third party mapping software, specifically Google Maps.

10. Survey mapper – post a survey to a public Web site, gather responses and store or map them (could be included in (3)).

The use of portlets to control the above steps in a workflow is illustrated in Figure 1. Portlets are an important interface because they facilitate user input to the workflow.

2.9 SUM Diagrams

See Figures 2, 3, 4, 5, 6, 7. These diagrams are described further in relation to a specific example later on in the next section.

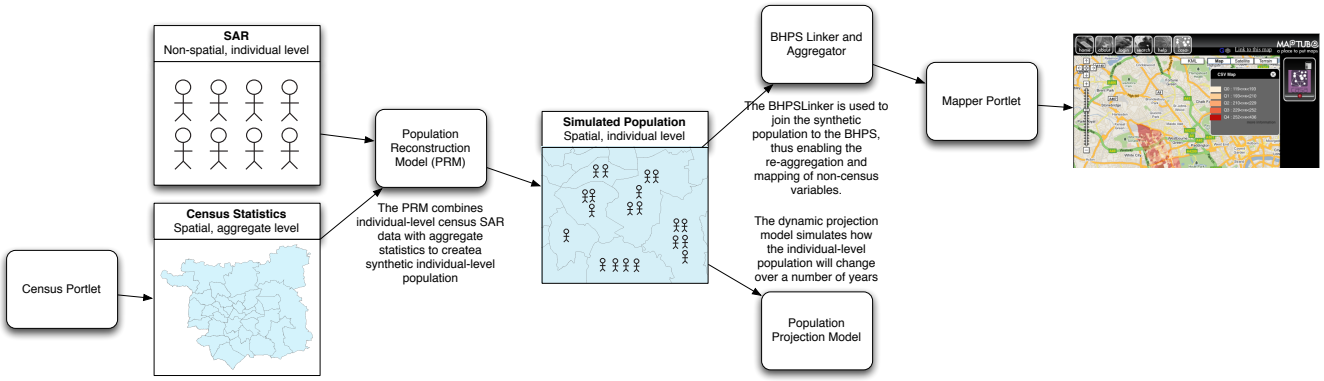

Figure 1: Overview of Simulation Portlets

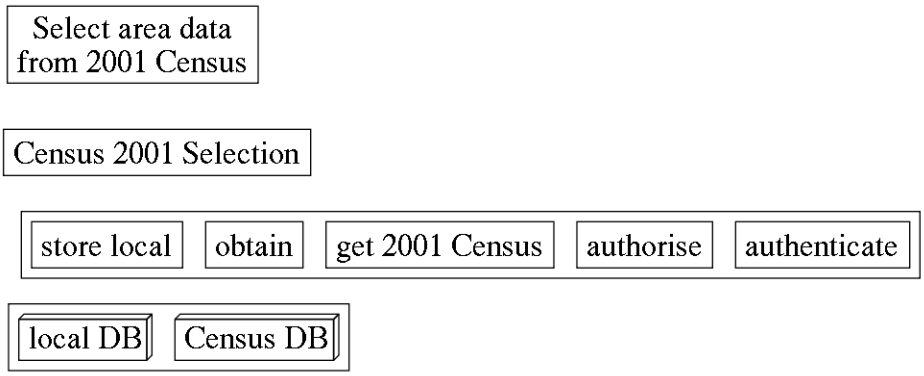

Figure 2: Selection from 2001 Census

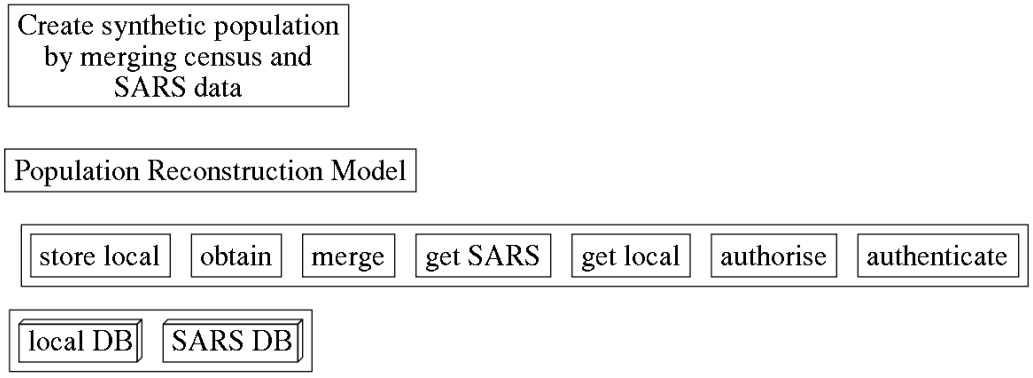

Figure 3: Population Re-construction Model

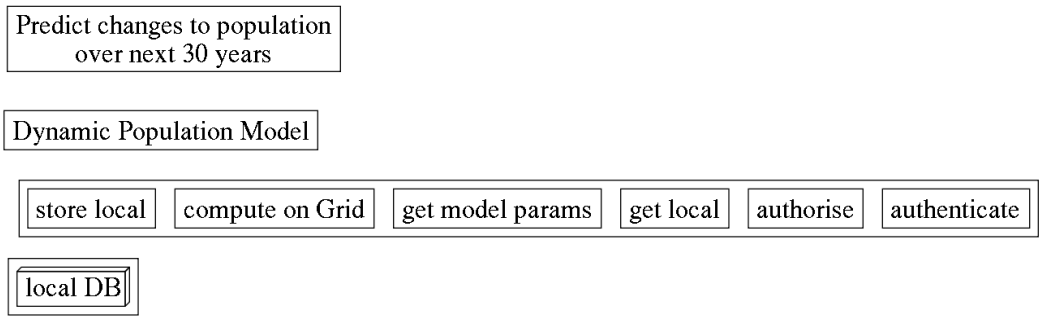

Figure 4: Dynamic Population Model

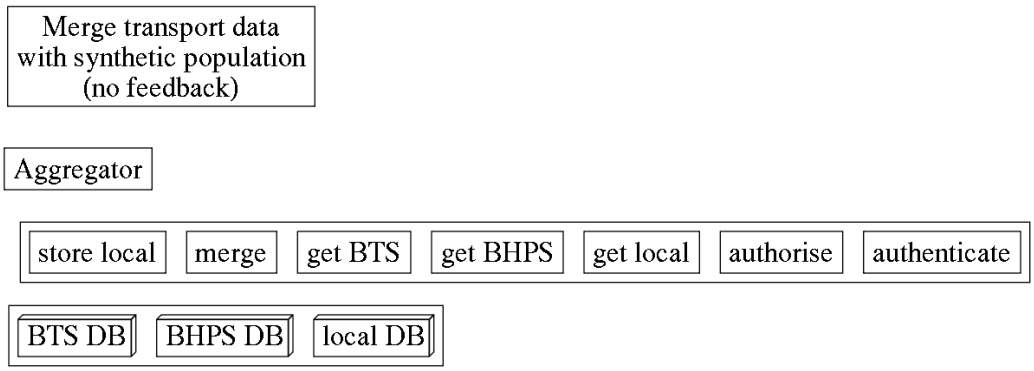

Figure 5: Data Aggregator

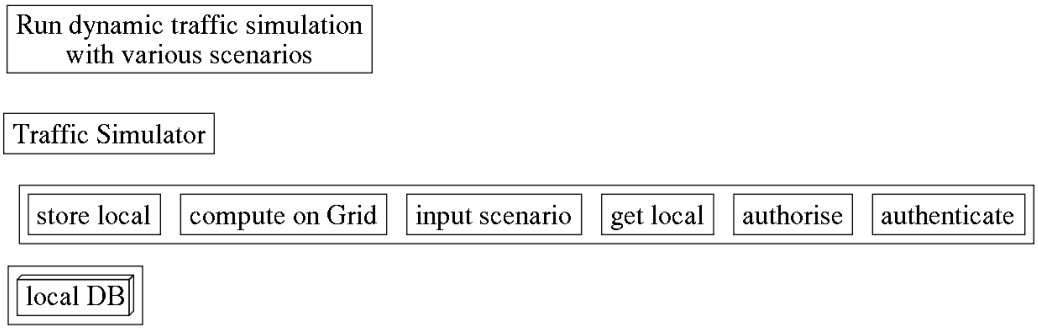

Figure 6: Traffic Simulator

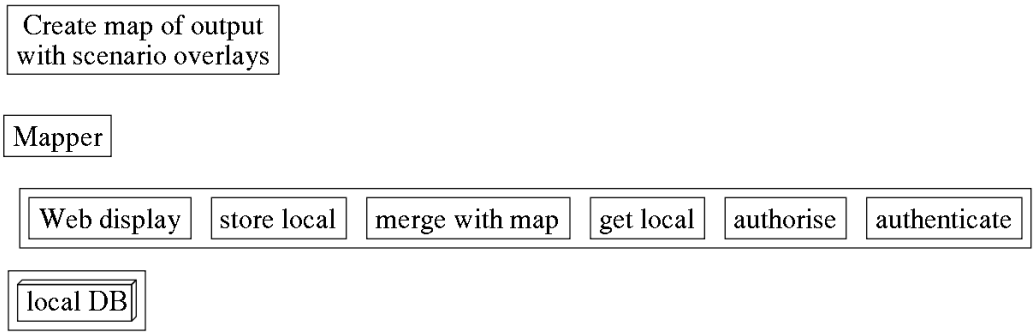

Figure 7: Mapping of Outputs

2.10 Usage Scenarios

The first scenario is from MoSeS and is a close fit to the business process described above. It posed a research question: *Linus is a geographer who is currently collaborating with a Primary Care Trust health and a social care analyst. They want to produce a report containing a map of current and forecast change in limiting long term illness distribution in Leeds at a health area level.* Note there is a separate MoSeS SUM document.

1. Login – Linus activates his login into NeISS;
2. Set-up and initialisation – Linus needs to action NeISS. This can be done by specifying the random seed;
3. Generate new demographic model data – new dynamic simulations and outputs are generated, e.g. new re-constructed populations for 2001;
4. Generate visualisations of the demographic data – generate visualisations of the demographic data at particular times and revealing changes over time;
5. Access stored demographic models – Linus can re-use or adapt the results generated and even download.

A transport modelling scenario has been elaborated within the NeISS project which further stretches the business model. This would be applicable initially to an individual medium sized city like Leeds or Bristol. It critically requires access to a number of, potentially confidential, data sources from which items need to be combined into synthetic data sets applicable to the time ranges of interest. This work is to be carried out using the DAMES services.

1. Crowd sourcing – pose questions and collect data using the SurveyMapper tool from CASA or a variant, e.g. how do you travel to work?
2. Access baseline demographic data for a city, e.g. from 2001 Census. This requires role based authorisation and access to the data server over some secure API. Select subset by area (input area code, name with dis-ambiguation, or click on a map). See Figure 2.

3. Population Re-construction Model (PRM) combines individual level Census Samples of Anonymised Records (SAR) data with aggregate statistics to create a synthetic individual level population. Iterate changing parameters until the required resolution is achieved. See Figure 3.

4. Dynamic Population Model (DPM) – starting from the PRM, use models of population dynamics to re-construct populations for future dates of interest. Access data from the dynamic model to look at the demographics of the city in 5, 10 or 15 years. See Figure 4.

5. Traffic generation model – aggregate population model with some kind of transport demand data. This data can be sourced from the BHPS [1]. Behavioural models may be used to identify different rates of making trips for different household types in different activity categories, e.g. work, shop, education, leisure. This could feed into a mapping portlet for display. See Figure 5.

6. Run a dynamic simulation model to allocate traffic to a road network. Potentially other modes such as train, bus, cycle, walk, etc. are also important. This uses spatial interaction models which are encapsulated as Web services. Diverse scenarios, e.g. road charging, congestion charging, bus lanes, are tested. Output indicators include, number of road accidents, average journey times, pollution levels, etc. See Figure 6.

7. Visualisation – using MapTube, e.g. to show the public how their individual decisions have an impact on the wider community. Outputs include maps and charts relating to the performance indicators, e.g. maps by neighbourhood, charts for the city, etc. activity and interaction patterns, e.g. work, shop, etc. See Figure 7.

8. Export as a workow to allow other users to re-run the previous steps with dfferent scenario data and visualise results without using all the portlet interfaces.

9. Take the simulation results from years 0, 5, 10, 15 and "publish" them to a Web site so that planners or members of the public can see what is being predicted.

10. Invite local residents to view the simulation outcomes then publish a survey via the SurveyMapper tool on what they think could be improved. For example, can we persuade some residents that it would be a good idea to leave their cars at home and get the bus to work on certain days?

11. Based on the results of the survey, we re-run an entire set of simulations with different traffic demand, e.g. assuming that some fraction of the population is willing to leave their cars at home on certain days. Plug in the new numbers and re-run the workflow.

12. Refresh the simulation with data from other sources, such as National Transport Survey 2002-8, British Social Attitudes Survey or Scottish Household Survey 2009?

2.11 Applicability

The NeISS project aims to deploy and demonstrate a generic e-Infrastructure for social simulation. Some areas of application include land use modelling, e-health and demographic planning.

[1] University of Essex. Institute for Social and Economic Research, British Household Panel Survey: Waves 1-18, 1991-2009 [computer file]. 7th Edition. Colchester, Essex: UK Data Archive [distributor], July 2010. SN: 5151.

2.12 Functionality

Based on the above process requirements and usage scenarios we have identified the following functional requirements.

- Authentication and role based authorisation – security is omni-present and some datasets are confidential and available only for specific purposes to named sets of users. This affects how datasets can be used severally or jointly. Functionality provided by this architecture was initially provided through the authentication and authorisation mechanisms employed by both the JSR-168 compliant portlet container (in the case of earlier demonstrators, GridSphere or LifeRay) and the NGS SARoNGS and VOMS services. Security is very important for NeISS as it incorporates confidential data (health and census data) which in NeISS will be accessed via services developed in the DAMES project. G-R-Toolkit can be used to wrap services and enable access only to authenticated users, see Figure 10.
- Session management – all objects created during a scenario run must be tagged with a personal or session id. This will facilitate polling the status of long running processes or saving and re-using data. It will also avoid conflicts when multiple users access NEISS. The session id is in the form of either a unique string or a SAML assertion.
- Data access and management – DAMES will provide user interfaces for selecting and manipulating data. Ideally, this would access the source data, using role based security. It is more likely however that access can be provided either in the form of approved and pre-prepared data extracts, or by users uploading their own data.
- Discovery and metadata – datasets are typically employed in a spreadsheet (CSV) format. It would be useful if there were further metadata describing the layout and content of these datasets and how then can be used or combined. In particular a method is needed to associate geo-spatial parameters with other data so that it can be displayed using mapping tools.
- Computationally intensive demographic and forecasting models – will use parallel processing on Grid resources. Examples include the MoSeS forecasting module and MoSeS demographic module.
- Data transfer service – to upload, download or transfer large quantities of data between resources. This may use the DataMINX service being developed for the NGS.
- Virtualised storage resources – Resulting datasets will need to be stored for use by collaborating researchers, further manipulated, updated or merged and exported as input to computational services. Data will need to be tagged with user's id. Output data will also need to be managed and stored, usefully with provenance information. Note that confidentiality issues must be addressed as noted above. Technologies of interest include Storage Resource Broker (SRB) or iRODS and DropBox in addition to DataMINX. Examples include MoSeS archive store, MoSeS forecasting store and MoSeS virtual population.
- Data fusion – DAMES will provide data fusion tools to aggregate synthetic population models with other data parameters. Initially a simple random assignment method is used.
- Portlets (JSR-168 and JSR-286) form the user interface to the underlying SOAP Web services of the project – examples include DAMES (GEMEDE, etc.), MoSeS Archiving portlet, MoSeS

Mapping portlet, MoSeS Charting portlet, MoSeS Analysis portlet, MoSeS Selection portlet. Portlets should be capable of being re-used in a variety of frameworks, specifically LifeRay and Sakai in this project. Both JSR-168 and JSR-286 are being used, the latter specifically for inter-portlets communication.

- Workflow construction, management and execution using distributed services with appropriate security – Taverna is used as the workflow enactment engine. The method of composition is not yet fully determined, but could be by automatic capture of metadata from the portlets as a scenario is completed. Taverna re-runs the underlying components as SOAP services, e.g. for a health service or transport planner.
- Maps, with associated workflow, metadata and survey results can be uploaded as "packs" to myExperiment for sharing with other researchers. myExperiment will act as a repository of work done using NeISS.
- Branding of Sakai work sites and myExperiment groups to be used together for NeISS work. Connect them using social networking technology, i.e. managed user and group lists.
- Presentation of scenario outcomes on a public Web site. This should also have functionality for surveys, e.g. using MapTube and SurveyMapper. Control how the map is displayed, e.g. data range and areas.
- Overlay maps to compare scenarios or different kinds of data.

2.13 Structure and Arrangement

Social simulation involves the creation and integration of digital artefacts across a varied set of activities in the research lifecycle. These artefacts may include primary datasets, simulation code, tools for analysis and visualisation of scenario outputs, and archives for deposition of results from simulation experiments or scenarios, see Figure 8. To accelerate uptake, researchers urgently require improved means for documentation, publication and re-use of simulations and their component parts. Social simulation can be computationally demanding, making it necessary to access compute resources at institutional, national or even the international level. Where researchers are interested in global social phenomena, collaboration adds to the complexity of the research process.

The NeISS architecture consists of four interacting layers, see Figure 9. The first is a services layer comprising the fundamental components on which simulation depends. These are data access, modelling and analysis tools, plus visualisation. The second is a composition layer in which individual services are composed into workflows and then published as coherent entities. In combination, layers one and two provide the basis for the integrated support of the social simulation research lifecycle. Above the composition layer sits an architecture layer. This provides tools and methods needed to provide portal access to simulation services and workflows and to combine these into domain specific exemplars (see Use Cases). In the deployment layer, these exemplars will be made available to users.

Data and computation is supported through existing e-infrastructures, including resources hosted at STFC Daresbury, Universities of Leeds and Glasgow, White Rose Grid, North West Grid and NGS with whom NCeSS partners have existing agreements for use.

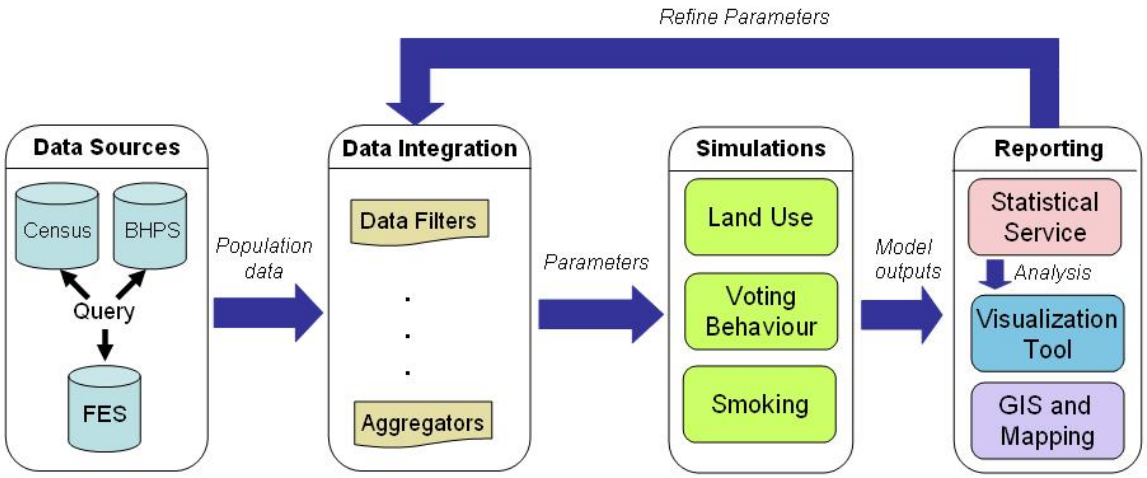

Figure 8: Overview of NeISS Workflow

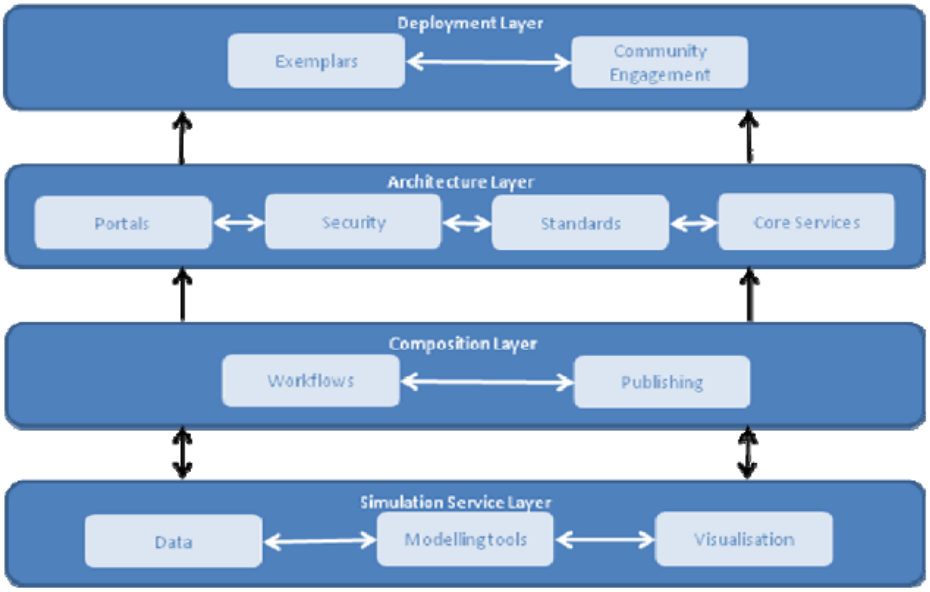

Figure 9: Layers in NeISS Architecture

2.14 Applicable Standards

This information will be maintained and kept up to date on the eReSS Wiki: `https://www.confluence.hull.ac.uk/display/context/National+e-Infrastructure+for+Social+Simulation`.

The proposed standards to be used in NeISS include the following.

Java: JSR-168, JSR-170, JSR-286, WSRP

OGF: JSDL, SAGA

Security: Shibboleth, SAML, X.509, XACML, TLS

Services: WS-I

Information: RDF, OWL, SPARQL

Social: FOAF, SIOC, OpenSocial

Web: HTML-4, Accessibility

2.15 Design Decisions and Tradeoffs

This is probably covered by above text as services are derived from previous projects as already described. Design decisions relating to security around a workflow accessing distributed services still have to be made.

2.16 Implementation Guidance and Dependencies

Some issues that may affect the implementation of applications that use the Service Usage Model, such as the organisation, performance, behaviours, representations and policies are noted here.

2.16.1 SARoNGS and CTS

Because of the nature of the CTS and how it interacts with the UK Federation it would be practically impossible to have a test or development service that would be of much use. NeISS has however experimented with a local version running on a server in Glasgow which has enabled testing of variants of the Eduserve person profile.

There are some current known issues with the CTS service as implemented for the NGS. These are as described in an NGS Operations e-mail on 21/6/2010. This means that the CTS is not currently a production quality service.

1) Using a VO other than a default listed one doesn't work. The system seems to be unable to pull down the correct information from the VOMS server to know how to make a request for an attribute certificate for a VO not configured ahead of time. For most users, this won't be an issue.

2) A way for users to get in and manage their NGS accounts is required. There's currently a Web page at `https://cts.ngs.ac.uk/Login/RegNGS.pl` that allows users to register for an NGS account using their SARoNGS identity. It seems that a copy and very minor tweak to this file would allow SARoNGS users the same access to their account information as users with a certificate in their browser.

3) Forwarding user names and passwords is currently done over plain HTTP. There's already a ticket open in the services helpdesk about this.

4) An API is required that would allow us to have a single page that will check if a user is registered on the NGS via SARoNGS, and possibly via regular certificates as well, and present them with links as appropriate; how to register (using either method, maybe) or links to the UAS interface that work for them etc. The current API sits at `https://cts.ngs.ac.uk/API`.

2.16.2 Access to Data

It seems highly unlikely that we shall gain direct access to data services, e.g. for Census or BHPS data. It may be that a subset of the appropriate data will have to be downloaded and hosted internally by the project to show somewhat more than a "proof of concept" service. We understand that it will be possible to have the data on the system in such a way that users are running jobs on it and reviewing results – so long as they can't access the original micro-data.

2.16.3 Access to Resources

Using SARoNGS currently only generates a "low security" user certificate which can give limited access to NGS and other resources. This is being investigated as noted above.

2.16.4 Workflow Security

Currently it is assumed that the entire workflow is executed within a single security domain. In this model the workflow enactment engine is accessed as a remotely hosted service which authenticates the user. Authentication and authorisation are then done once only. In fact currently the workflow is only accessible from the portal as a "back end service" in the same way that MySQL is used to store the internal portal data. We would like to investigate workflows composed of autonomous services, a much harder problem.

The conceptual workflow for transfer of security messages in G-R-Toolkit is illustrated in Figure 10. This uses the NGS MyProxy service for certificate handling in a portal, possibly following certificate generation with Shibboleth and the SARoNGS CTS service. Individual services are then protected using access control lists and TLS handshaking via the GridSite Apache module.

2.17 Known Uses

Particular exemplar application areas include the following.

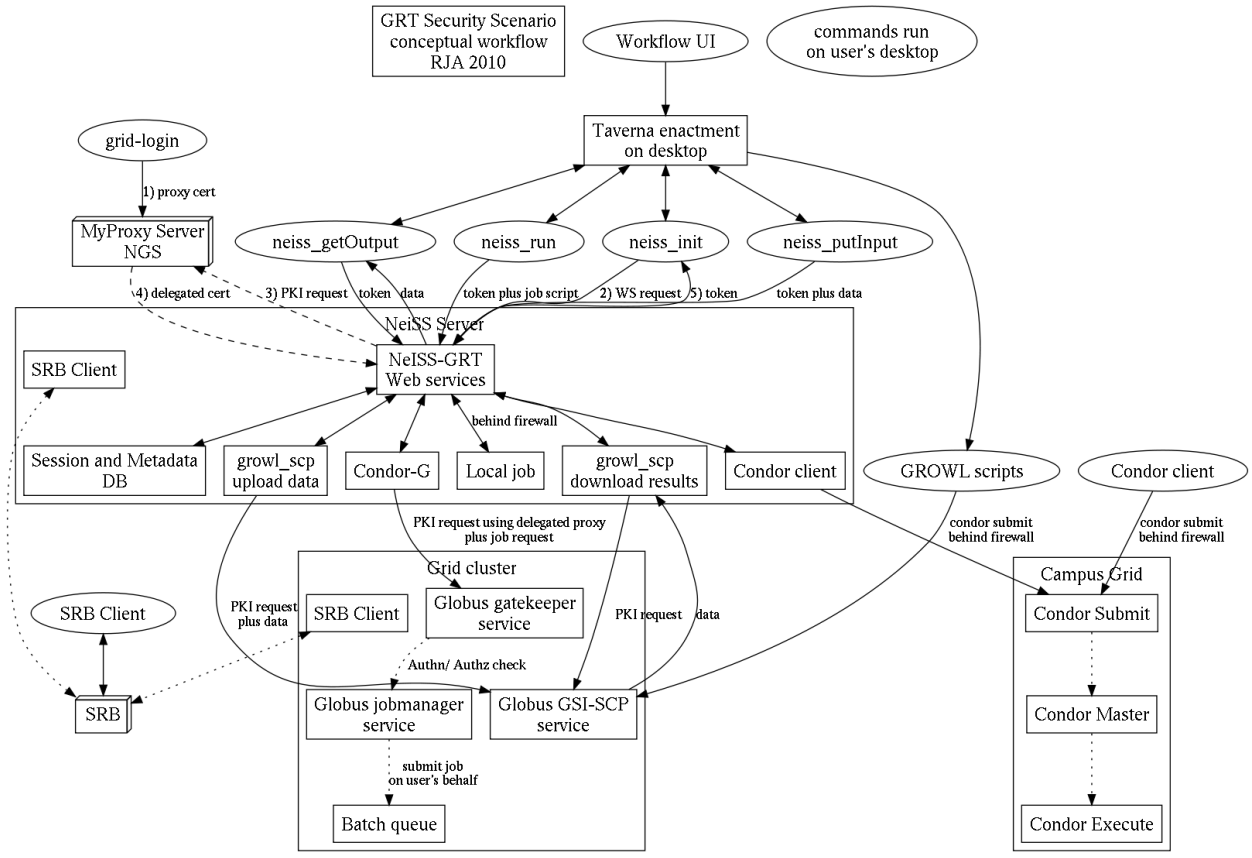

Figure 10: Security in G-R-Toolkit

SimCity Transport: – Mark Birkin. As described above.

Epidemiology: – Dexter Canoy.

Ageing and Inequality: – Paul Lambert.

Credit Crunch: – Richard Milton.

SimTube: – Andy Hudson-Smith. 3D walk through SecondLife type exhibition space to show outcomes of the project for outreach purposes.

Each exemplar owner has described the kind of research questions they could address using the NeISS Infrastructure. The exemplars are written up separately.

2.18 Data Sources Used

Some or all of the following might be accessed and data combined. Note the likely restrictions discussed above.

- UK Census;
- British Household Panel Survey;
- National Transport Survey 2002-6;
- British Social Attitudes Survey;
- Scottish Household Survey 2009.

2.19 Related SUMs

UK Federation Access Management `https://e-framework.usq.edu.au/users/wiki/DevelopmentSUMUKFed`.

Grid Enabling MIMAS Services (GEMS) `https://e-framework.usq.edu.au/users/wiki/DevelopmentSUMgems`.

MoSeS (Modelling and Simulation for e-Social Science) `https://e-framework.usq.edu.au/users/wiki/DevelopmentSUMmoses`.

2.20 Services Used

The following are either individual services or service frameworks.

- LifeRay;
- Sakai;

- myExperiment;
- Taverna;
- Shibboleth, SARoNGS, CTS;
- DAMES;
- PRM;
- DPM;
- MapTube;
- SurveyMapper;

2.21 CORE SUMs Used

List the names and versions of all Commonly Recurring (CORE) SUMs that this Service Usage Model is based on.

No relevant ones found on the e-Framework Web site or TRAC Wiki.

2.22 References

See Web links in Notation Section 2.6.

2.23 Terms

See Notation Section 2.6.

3 Contributors

Contributors to this document include the following. Mark Birkin, Neil Chue-Hong, Andy Hudson-Smith, Paul Lambert, Bill Lin, Nick Malleson, David Meredith, Alexandra Nenadic, Andy Turner, John Watt.