Security-oriented e-Infrastructures for Social Simulation

J. Watt, S. McCafferty, R.O.Sinnott
National e-Science Centre,University of Glasgow
j.watt@nesc.gla.ac.uk

Introduction

The JISC-funded National e-Infrastructure for Social Simulation (NeISS) [1] project aims to develop and provide new services to social scientists and public/private sector policymakers interested in "what-if" questions that have an impact upon society and can be tackled through social simulation. For the first what-if question, a traffic simulation modelling how congestion will affect routes within a city or region projected across a time-span of decades has been identified. This paper describes the work that has been done in implementing a secure, user-oriented environment that provides seamless access to relevant nationally significant data sets such as the 2001 Census and demographic transition statistics from the British Household Panel Survey (BHPS) [2], and a Population Reconstruction Model (PRM) simulator, which simulates a population of individuals or households based upon these data sets.

Traffic Simulator and Associated e-Infrastructure

The PRM simulator produces synthetic populations at the individual and household level, created from the Census Sample of Anonymised Records (SARS – http://www.ccsr.ac.uk/sars/). These populations are highly representative of the actual population, as they are created from an anonymised subset of the raw census responses at the individual level. The Census data provided by the NeSC Grid Service described below is the CAS (Census Area Statistics), which provides both univariate and key statistics. These are granulated at the Census Output Area level (or equivalent small region), thus ensuring that any real demographic trends are magnified. As part of the simulation models, these initial populations undergo transformation to forecast what the final population will look like in the future. The transition rates themselves are extracted from the longitudinal BHPS for effects such as ageing, fertility, mortality and household formation [3]. The information produced by the PRM forecaster is fed into traffic simulators to provide an accurate reflection of future population trends and the impact upon transport.

The interface for the PRM simulator is a web portal populated with JSR-168 compliant portlets. The most recent standard (JSR-286) was avoided as this presented problems when porting to other portal frameworks to be exploited in the project such as Sakai. To enable data linkage across portlets interacting with remote services, use was made of the standard PortletSession ID scoped across all portlets for a given user session. By passing this sessionID to remote services, a reliable link between data generated by a Census data Grid service and the data required by the PRM may be established.

The portal utilises the Shibboleth federated authentication mechanism for user identification. A modified version of the SPAM-GP [4] CCP module allows all UK Federation member IdPs to log into the portal from their home institutions – with the eduPerson attribute 'eduPersonTargetedId' used as the sole identifying user credential. Users then personalise their accounts themselves as UK Federation IdPs tend not to release any disclosing information by default. Portal-centric Role Based Access Control may be enforced by the portal administrator by issuing roles to individual users and then restricting access to subsets of the portal only to users asserting these roles.

A simplistic (but user-oriented) Census data portlet that allows a range of variables to be selected based upon the specific scenario in mind. This portlet interacts with a GT4-based Grid service which allows users to retrieve information from the database indirectly. That is, through providing this Grid Service with the desired data query (from the portlet), along with the aforementioned sessionID for identifiability a resultant data set is created as a CSV file containing the requested data, and this is returned as a URL for downloading from a secure web server. This accessor-oriented service model allows other infrastructures to interface with the database and retrieve information through use of URLs. Thus for example it is planned as part of the NeISS project to exploit Taverna to co-ordinate the interactions between services (simulation and data services)..The data created by the Census Data Portlet can be sent to the CASA MapTube Web Service, which allows basic rendering of supplied Census data onto UK maps. The PRM simulator data feeds into the Dynamic Population Simulator which links the constructed populations with the BHPS data stored locally. Figure 1 shows some of the interactions between these components, and Figure 2 shows some of the tools currently deployed on the NeISS portal.

An additional layer of security for accessing the Grid Service is added utilising a further tool from the SPAM-GP suite, namely the Attribute Certificate Portlet (ACP). Currently, access to the service is through a proxy certificate generated from a NeISS-specific certification authority (CA). This security is further enforced through firewall locking to known requesting resources. However a more flexible and powerful approach is through use of the ACP, where a PERMIS Policy Enforcement Point (PEP) has been loaded into the Globus infrastructure, which denies access to resource requests except where a valid X.509 Attribute Certificate exists for valid (authenticated) users. This raises an important issue since access to the portal through Shibboleth (and signed SAML assertions) and the de facto Grid security model of authentication through X509 based Public Key Infrastructure and use of proxy credentials needs to be aligned. Whilst the stated goal of most e-Science/e-Research projects is hiding as much user interaction with underlying security/middleware technologies as possible, the back-end solutions required to invoke PERMIS protected GT4 services (i.e. authorisation) depends firstly upon authentication through Globus Grid Security Infrastructure (GSI). As a result other portal-Grid models, for example exploiting a single server certificate to interact with Grid resources would not work (since every user who accessed the portal would be authenticated/authorised to invoke the Grid service if the same server certificate were used). This is currently tackled through having the NeISS CA and use of tools such as MyProxy and VOMS. However in this case it is necessary for the user to undertake an additional authentication step to create their proxy certificates within the portal (in addition to providing their Shibboleth identity). One potential solution to this is provided by the SARoNGS [5] infrastructure, which provides a means of generating low-assurance proxy certificates from a SAML assertion from a trusted Identity Provider, and using them to run jobs from the NGS portal. The SARoNGS infrastructure can also be called from external portals, with the infrastructure returning MyProxy credentials to the requesting portal which can be used to automatically download short-lived proxy certificates. Issues such as these will be tackled when the PRM/traffic simulator are deployed on the NGS. A second solution which has been adopted is to use a targeted IdP at NeSC Glasgow where encrypted username/password combinations for MyProxy are used to create an X509 proxy credential as part of the process of establishing the portal session. This solution is limited in that it demands the identity provider to offer such information (and this will typically not be the case for arbitrary IdPs existing across the UK federation).

Furthermore data curation and auditing are critical factors in NeISS. A reliable archive of the calls to the database, the data extracted, and by which external service needs to be captured, stored and subsequently managed. By default, data generated by queries through the Census Grid service are stored in the backend database as per-query tables. The sessionId used to link the data in the portlets is also used as an identifier in the database to enable a link between the data requested and the subsequent information generated by the modeller. Since the sessionId is an opaque identifier which reveals no direct information on the requester, the portlet will be responsible for logging the mapping between sessionId and portal account.

Conclusions

The security model developed at NeSC over several projects (SEE-GEO, DAMES) has been refined for use in a real-life social simulation scenario within the NeISS project. A JSR-168 portlet has been developed which interfaces with a GT4 data service to allow secure and accountable access to sensitive Census Data tables. This portlet is fully portable to any remote portal as all the database calls are done through the data service, using a proxy certificate from a custom CA. Using the portal sessionId, this data can be linked directly to the PRM tool to allow seamless merging of data and simulation. The data service can also be linked to workflow technologies like Taverna, enabling a new method of information extraction and processing. Future work includes hardening of the archive features, and integration with large-scale resources such as the NGS through novel certificate management. If accepted, the full paper will expand upon these issues and describe detailed traffic-related simulation scenarios supported through the NeISS project.

References

[1] National e-Infrastructure for Social Simulation (http://www.neiss.org.uk)

[2] UK Data Archive British household panel survey. See http://www.data-archive.ac.uk/

[3] M. Birkin et al. “The Elements of a Computational Infrastructure for Social Simulation” Philosophical Transactions A, Royal Society. Proc. Of AHM09

[4] J. Watt et al. “Tool support for security-oriented virtual research collaborations”. In Proc.
IEEE int. symp. on parallel and distributed processing with applications, pp. 419–
424. (doi:10.1109/ISPA.2009.49).

[5] X. Wang et al. “Shibboleth Access for Resources on the National Grid Service” Journal of Information Assurance and Security 5 (2010) 293-300

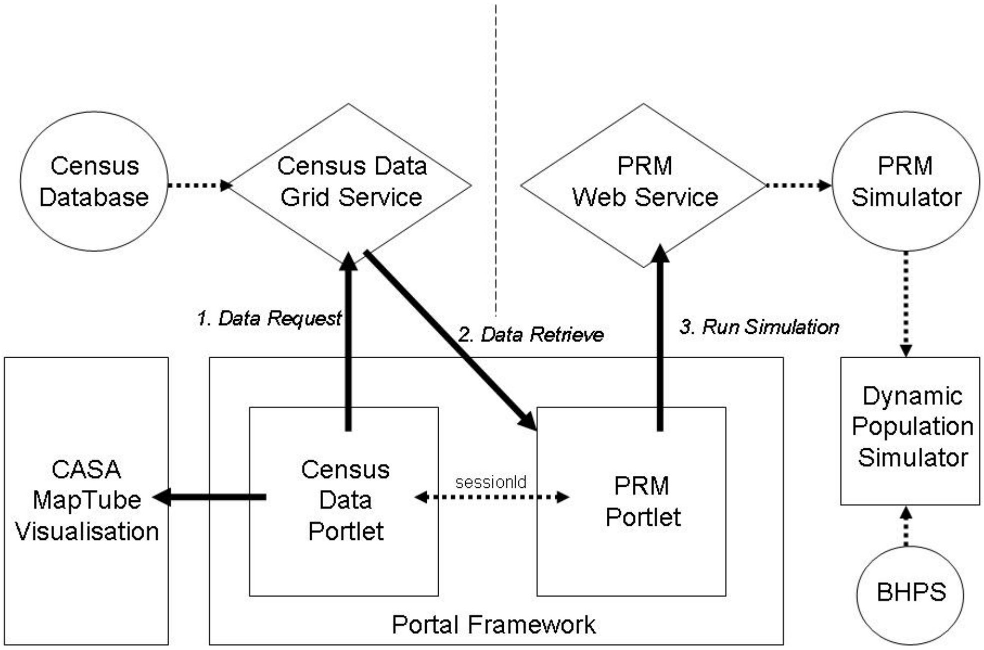
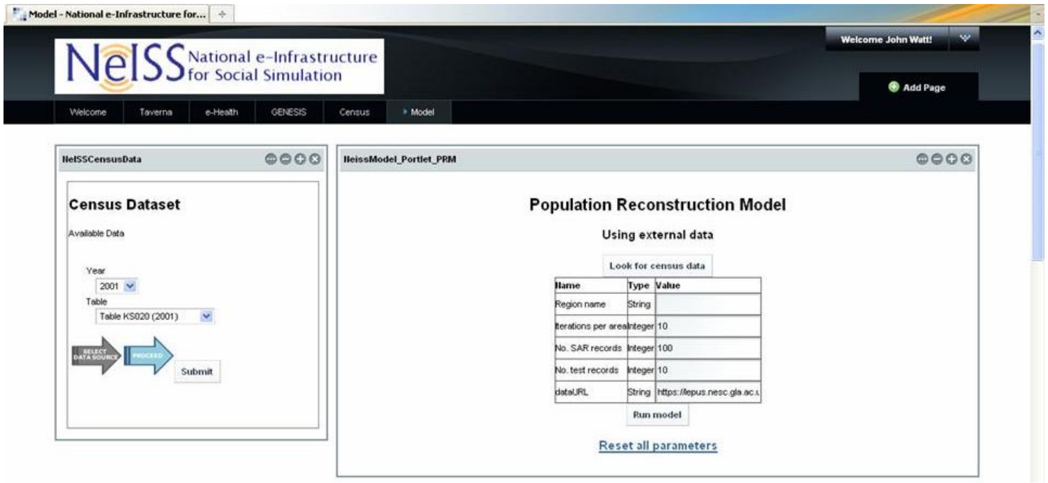


Figure 1: Schematic of the interaction between the Census Data Service and the PRM Web Service, the CASA drawMap functions, and the final Dynamic Population Simulator.



Census Dataset

Available Data

Year: 2001

Table: Table KS020 (2001)

Population Reconstruction Model

Using external data

Name	Type	Value
Region name	String	
Iterations per area	Integer	10
No. SAR records	Integer	100
No. test records	Integer	10
dataURL	String	https://lepus.nesc.gla.ac.u

Figure 2: The NeISS Portal showing the Census Data Portlet and the PRM Portlet