

# **Proposal: e-Infrastructure for the Social Sciences**

## **National Centre for e-Social Science**

### **February 2006**

#### **1. Introduction**

The 2003 ESRC-commissioned e-Social Science Scoping Studies, the NCeSS Node Commissioning Panel, the Agenda Setting Workshops (ASWs) and the external e-Social Science Review of 2005 have confirmed that Grid-enabled datasets, services and tools are key enablers for the wider take-up of e-Social Science. In particular, the Review concluded that it was important to create a coherent e-Infrastructure that would provide a platform for disseminating the benefits of e-Social Science to the wider social science research community, leverage existing e-Social Science and e-Science investments, and ensure the usability and sustainability of middleware, services and tools. This proposal sets out a plan for an initial build of this e-Infrastructure.

The project will build on the work of the NCeSS programme members and other UK social science stakeholders such as the major data providers. NCeSS is composed of seven 'Nodes' carrying out substantive and methodological social science research over three years, twelve 'small grant' projects, mostly of one year duration and a Hub at Manchester University/UKDA responsible for overall coordination. The ESRC-JISC funded ReDReSS project is responsible for some additional aspects of e-Social Science training and awareness raising, and is also working on user portals. All will contribute to building the e-Infrastructure. Collectively, the project team brings together the key stakeholders and researchers in UK e-Social Science.

#### **2. e-Infrastructure Goals**

We will build an e-Infrastructure on the UK National Grid Service (NGS) to provide integrated access to a variety of resources for social science research, including datasets, tools, services and easy-to-use user environments. In so doing, it will deliver important benefits for UK e-Social Science, and also address the priorities identified in the Review. Specifically, the e-Infrastructure will:

- make available new, powerful, easy-to-use research tools and services which will reduce the effort needed to integrate and re-use datasets, and simplify the way that research is currently carried out;
- make available research demonstrators to illustrate the capacities of the new research tools and the future potential of e-Social Science;
- enhance our understanding of issues around resource discovery, data access, security and usability by providing a testbed for the development of metadata and service registries, tools for user authorisation and authentication, and user portals and collaborative virtual user environments;
- increase e-Social Science capacity building by training end-users and developing technical expertise;
- lay the foundations for an integrated strategy for the future development and support of e-social science infrastructure and services, and produce a road-map that identifies the human, technical and financial resources required to pursue the strategy;
- leverage the infrastructure investment being made by UK e-Science core programme and JISC for the benefit of the Social Sciences;
- provide exemplars of best practice and ready-made solutions to common requirements to support NGS, JISC and OMII-UK in ensuring that tool developers and service providers are led towards compatible and effective software and practices;
- promote synergies across NCeSS and other ESRC investments, co-ordinate activities, encourage mutual support and identify areas in which to promote the benefits of common policies and technology standards.

The key elements in building the e-Infrastructure are as follows:

- engaging with the social science community to ensure that the project is driven by research needs and, specifically, to identify the most research-relevant resources, tools and services to incorporate into the e-Infrastructure (WP1);

- working with major ESRC/JISC funded data centres (ESDS, MIMAS, EDINA and UKDA) and non ESRC/JISC funded data holders (e.g., ONS, OS) to Grid-enable selected quantitative and qualitative datasets, define metadata and ontology standards, and build registries to support resource and service discovery (WP2);
- harnessing the collective efforts of the NCeSS Hub and Nodes to prepare selected Grid-enabled, tools, services, demonstrators and user portals for use by the wider social science research community (WP3, WP4, WP5);
- developing training materials and support programmes for end-users, including online documentation and tutorials, and hands-on workshops (WP6);
- working with the UK e-Science core programme (NGS, OMII-UK consortium, NGS, NDCC) and JISC to devise an e-Infrastructure development plan which will define technical standards and mechanisms to ensure long term sustainability of the e-Infrastructure (WP7).

In the following sections, we provide some examples of how social science research can benefit from e-Infrastructure. We then give an overview of the e-Infrastructure's components. The main tasks and contributions of the project partners are then defined in a series of eight workpackages (WPs). The proposal also outlines how the project will engage with stakeholders: ESRC/JISC- and non ESRC/JISC-funded data centres, the UK e-Science core programme (National e-Science Centre, National Centre for Text Mining, NGS) and JISC. The project will be managed and coordinated by the NCeSS Hub (WP8).

### 3. e-Infrastructure Benefits for Social Science Research

As documented in the first Annual NCeSS Report, investigations of the potential for the take-up of e-Social Science has uncovered three distinct communities – the 'early adopters' who are keen to push to the limit of what is possible; the 'interested' who will adopt new research tools and services if they believe these will provide simple ways of advancing their research; and the 'uncommitted' who have yet to appreciate the relevance of Grid technologies for their research. Those early adopters who have not been recruited into the NCeSS programme will demand research tools they can apply. We also need tools that will enable us to convert the interested into adopters and demonstrators which will convince the uncommitted to join the ranks of the interested.

In this section, we give some examples of how the early adopters and interested communities would be able to use e-Infrastructure to enhance their research and how we will engage with the uncommitted by providing demonstrators. The examples illustrate how social scientists can combine its component parts – i.e., datasets, services and tools – in flexible yet powerful ways to overcome various kinds of problem they face in pursuing their research.

**Early adopters:** Solving complex statistical modelling problems often involves multiple steps and iterations where the output of one step is used as the input in the next: 1) select data or subsets, maybe from more than one source; 2) merge, harvest or fuse; 3) input to model and analyse; 4) repeat previous cycle with new or different data; 5) repeat with different model assumptions, parameters and possibly data; 6) synthesise outputs from multiple models or world views; 7) load output into a different analysis or visualisation tool.

Managing these steps manually is potentially difficult, and performing the data integration and modelling using desktop PC tools may be very time consuming. For example, analysis of work histories requires many different sources of data which need to be reconciled and integrated in order to produce a coherent and contextualised life or work history and takes one week on a desk top PC in Stata. To then simultaneously analyse the data could take months on a desktop PC running serial SABRE<sup>1</sup> and over 10 years using Stata. This presents a major obstacle to research. Using Grid-enabled data, analysis and workflow tools, however, the researcher can compose complex sequences of steps using different computational and data resources, and execute them (semi-)automatically using powerful computers in a comparatively short time.

**Interested:** Schelling's model of segregation provides a way of exploring how the spatial distribution of society groups ('agents') responds to different assumptions about neighbour preferences. For example, a social scientist interested in racial segregation could use it to explore questions such as:

---

<sup>1</sup> [http://www.growl.org.uk/Papers/GROWL\\_AHM\\_paper.pdf](http://www.growl.org.uk/Papers/GROWL_AHM_paper.pdf)

- How sensitive the model is to the initial arrangements of agents.
- The relationship between the ratio of agents of different ‘races’ and the final segregation ratio.

To do this, the researcher might need to run simulations for 50 different values of the initial race ratio and 100 different initial agents’ arrangements. Using a desktop PC, each simulation run would take around one minute, so 5000 simulations will take about one week, which may well make it impractical. If the researcher has access to Grid computational resources, however, the same number of simulations can be run in about five minutes, opening up new possibilities for the researcher.

**Uncommitted:** We will deploy a selection of demonstrators from those being developed within the Nodes (e.g. GeoVue Pollution and Virtual Cities 3D visualisation tools), small grants, and those already developed by the e-Social Science Pilot Demonstrator Projects (PDPs).

#### 4. e-Infrastructure Architecture Overview



Figure 1: The main components of the e-Infrastructure.

**Portals** provide an integrated, single point of access to e-Infrastructure resources through a familiar and simple-to-use web-style user interface which hides the underlying complexity. Users can authenticate themselves, discover resources (data, tools and services) and create their own ‘workflows’ from tools and services to carry out analysis. A tool or service is mapped to a Java portlet for insertion into one or more portal frameworks. Each Node can deploy its own preferred framework and choose from a repository of portlets depending on its user requirements. NCeSS will also host fully-functional portals with both collaboration tools and service portlets for access to Grid-enabled datasets and tools, including tools for

generic tasks such as resource discovery, workflow composition, data mining and visualisation. These portals will be designed to be easily usable by a wide range of users. **Collaboration virtual environments** (such as Sakai) help distant researchers: share in the research tasks, run project meetings, discuss results, and work up presentations and papers.

A range of **research tools** and **demonstrators** will be selected from those being developed within the NCeSS programme, with selection criteria informed by ongoing consultations with the wider social science research community. Agenda Setting Workshops have already enabled us to identify one priority area: simulation and modelling. Building on the work of the MoSeS and Policy Grid Nodes, simulation and modelling tools will be deployed which will enable researchers to run their own simulations, visualise and analyse results, and archive them for future comparison and re-use. This will facilitate development and sharing of social simulation resources within the UK social science community, encourage cooperation between model developers and researchers, and help foster the adoption of simulation as a research method in the social sciences, issues highlighted by the Agenda Setting Workshop (October, 2005) on e-Infrastructures for Social Simulation.

A set of **common tools and services** (e.g., workflow, visualisation) will also be made available. These will include tools and services developed within the e-Science community (e.g., Kepler<sup>2</sup>). We will also investigate how the project can build upon e-Social Science PDP demonstrators such as ConvertGrid. Our approach to e-Infrastructure building does not preclude local community-level deployments of tools and services. Indeed, the NCeSS programme is uncovering requirements for usability, control and trust in qualitative research that may be more appropriately served through localised provision of tools and data. For example, there may be issues with the distribution of sensitive video data to (and via) un-trusted third parties. Development of desktop tools and community-oriented service deployments is part of the ongoing NCeSS e-Infrastructure strategy, particularly in supporting qualitative research. The MiMeG and DReSS Nodes are building new tools and have a remit within their existing work programmes to capacity build and deploy to interested communities. MiMeG will be releasing webservices and tools that can be downloaded from a **community repository** and locally deployed later this year.

**GROWL**<sup>3</sup> (Grid Resources on a Workstation Library) is middleware that allows users to exploit Grid resources such as datasets while continuing to employ their favourite desktop analysis tools (e.g., Stata).

A range of **Grid-enabled datasets** will be made available (quantitative and qualitative), including datasets already Grid-enabled by PDPs (e.g., BHPS) or being Grid-enabled by MIMAS (Census 2001) and EDINA. Other datasets will be selected after consultation with data providers and researchers. Grid-enabling datasets can be done in two ways: 1) use OGSA-DAI on the NGS to connect to data resources held elsewhere; or 2) host the data in Oracle on the NGS. In either case, metadata is then be added to describe the data.

By defining a portable machine-processable semantics for resources, **metadata, ontologies and registries** facilitate resource discovery, enable tools to interact and complex workflows to be created. UK e-Science projects have used a family of XML-based technologies, most notably the Resource Description Framework (RDF), to provide a mechanism for representing resource metadata. Ontologies capture the meaning of metadata terms and their interrelationships. The Web Ontology Language (OWL) provides a vocabulary for describing classes of RDF resources and their properties (including relations between classes, cardinality, etc.). Registries provide ‘added value’ indexing schemas relevant to a particular community of use and by simplifying navigation of terms by enabling multiple schemas to be accessed from one view. Existing metadata infrastructure in use within the social sciences (such as the Data Documentation Initiative – DDI) as well as other activities within the UK e-Science core programme will help us develop provenance support. In the longer term, the contents of the metadata registry will be subject to discussion and eventual international agreement to permit sharing of resources world-wide.

## 5. Work Plan Outline

In this section we outline the main activities in building the e-Infrastructure as a series of workpackages, defining the partners involved and the resources required. Where applicable, the first named against each WP will be responsible for coordinating the activities of the other WP partners.

---

<sup>2</sup> <http://kepler-project.org/>

<sup>3</sup> [http://www.growl.org.uk/Papers/GROWL\\_paper\\_Final1.pdf](http://www.growl.org.uk/Papers/GROWL_paper_Final1.pdf)

**WP1 [4 person-months] Requirements gathering (Hub).** The objective of this WP is to ensure that the project is driven by the needs of social science researchers. It consists of two consultation activities which will be conducted iteratively over the course of the project, giving the opportunity to review progress and identify requirements as they emerge.

**WP1.1 [2 person-months] Datasets review (Hub).** The WP will conduct a consultation exercise with stakeholders to prioritise and select datasets to be Grid-enabled by the project.

We will consult with NCESS Nodes and small grant projects to identify dataset usage within the NCESS programme; the major data centres to identify patterns of dataset usage (quantitative and qualitative) within the wider social science community, to understand licensing issues and ensure complementarity with JISC funded Grid-enabling activities (current and planned); with experts such as Professor Elias; and (through the ASW programme) with the social science community to identify research drivers and ensure a fit with the ESRC's future data strategy plans.

We will use experience gained from previous (e.g., GEMEDA PDP, GEODE NCESS small grant) and current Grid-enabling projects (MIMAS and EDINA) to inform estimates of effort required for Grid-enabling.

Deliverable	Description	Due	Effort
D1.1.1	Prioritisation, costing and selection of datasets for Grid-enabling.	6, 12	2CC

**WP1.2 [2 person-months] Research drivers review (OeSS, Hub).** This WP will conduct a longitudinal review of research drivers for e-Infrastructure to establish and track priority areas. One instrument will be a Web-based survey of the UK social science community. In addition, there will be selective interviews with researchers who will have various degrees of awareness of e-Social Science. The WP will complement the Hub's existing intelligence activities (Agenda Setting Workshops and the Social Science Forum) with surveys and interviews with members of the UK social science community. Similar work is being undertaken by JISC in relation to UK e-Science and the WP will build on this and other studies.

The survey and additional information gathering will seek to identify who uses and does not use e-social science, and who is aware and unaware of basic concepts and tools. These patterns of awareness and use have implications for the sustainability of e-social science, and will help to inform policy and practice within NCESS. The findings will be incorporated into the report under WP7 below.

Deliverable	Description	Due	Effort
D1.2.1	Web-based survey of e-social science awareness.	8, 22	1.5
D1.2.2	Selected interviews with e-social science community.	22	0.5
D1.2.3	ASWs, Social Science Forum, etc.	Continuous	Hub resourced

**WP2 [22 person-months] Grid-enabling datasets (NCESS).** Guided by **WP1.1**, a number of datasets (quantitative and qualitative) will be selected for Grid-enabling over the course of the project. For example:

- The British Household Panel Survey (UKDA, ISER). As the BHPS is the most popular dataset distributed by the UKDA, its Grid-enabling is likely to be useful to a significant number of advanced users and to interest an even wider community to the potential of e-Social Science. This would build on the development work undertaken as part of the GEMEDA PDP.
- National and international time series (ESDS International @ MIMAS), such as the ONS Time Series, the IMF International Financial Statistics and the OECD Main Economic Indicators.
- Neighbourhood Statistics (ONS). This would build on the development work undertaken as part of the ConvertGrid PDP.
- Qualitative datasets: the project team will liaise with ESDS Qualidata over its plans to Grid-enable some of its datasets and determine how to take them forward in the context of this project.

Grid-enabling datasets can be done in two ways: 1) use OGSA-DAI on the NGS to connect to data resources held elsewhere; or 2) host the data in Oracle on the NGS. In either case, metadata is then added to describe the data.

**WP3 [60 person-months] Common infrastructure definition and development (Hub, Nodes).** This WP will develop the common components of the e-Infrastructure.

**WP3.1 [18 person-months] Metadata schema, registry specification and development (PolicyGrid, Hub, CQeSS).** This WP will focus on identifying the nature and scope of existing metadata schemas/ontologies, and a series of use cases for tasks such as discovery, data exchange and integration. Based on these, it will develop and deploy an initial flexible metadata model using RDF and OWL with a registry and appropriate tools taken from previous projects. A requirements gathering exercise will be conducted involving NCeSS programme members and other stakeholders such as the UKDA. The WP will also build upon the work of the 'Data chronicles' small project grant which is developing a pilot social science metadata repository.

A service registry will allow tools to publish and discover service information, allowing tools to interact and complex workflows to be created. UDDI provides a possible technical solution, e.g., using the open-source jUDDI implementation as investigated by the UK Grid Engineering Task Force (ETF), JISC work is comparing this with others, such as the Information Environment Service Registry (IESR).

For metadata to be widely adopted within e-Social Science, high quality tools are needed to allow social scientists to annotate their resources, and to query information managed by the various registries and services. The PolicyGrid Node is developing simple software tools that allow social scientists to create metadata as unobtrusively as possible, and to query such resources. These tools will be 'hardened' for wider use within the e-Social Science community.

Deliverable	Description	Due	Effort
D3.1.1	Metadata Requirements Definition.	6	2
D3.1.2	Metadata Registry.	9	2
D3.1.3	Populated Metadata Registry.	12	2
D3.1.4	UDDI Service Registry.	6	2
D3.1.5	Annotation Tools (datasets, services ...).	15	4
D3.1.6	Query & Presentation Tools.	15	4
D3.1.7	Documentation.	18	2

**WP3.2 [9 person-months] Authentication and authorization service (Hub).** As the JISC strategy of building a Shibboleth-based framework for authentication and authorisation unfolds during 2006, we expect to see a significant and growing proportion of UK social scientists working for institutions that operate Shibboleth Identity Providers. It is therefore important that the Web-accessible tools and datasets on which social scientists increasingly rely support Shibboleth-based authentication and authorisation schemes. Currently, many such tools and most publicly curated datasets of social science interest rely on the Athens access management system, and the community investment in Athens-based access control lists needs to be preserved. At the same time, there is a trend towards provisioning Web portals using Grid resources such as the National Grid Service, which relies on a Public Key Infrastructure. Faced with a period of churn in the authentication and authorisation domain, providers of services to e-Social Scientists urgently need a pragmatic strategy in order to both preserve investment and adapt to the emerging environment.

The outcomes of this WP will be (D3.2.1) an up-to-date review of the authentication and authorization landscape, including advice and recipes for making resources currently managed through Athens or the Grid Security Infrastructure accessible through Shibboleth, and (D3.2.2) the implementation of these recipes in Web portals that provide access to strategic tools and datasets.

The WP will follow developments in the Eduserv's Shibboleth-Athens gateway<sup>4</sup> as well as projects funded by JISC to Shibboleth-enable the NGS and its portal interface (ShibGrid and SHEBANGS<sup>5</sup>) and evaluate their outputs (expected between Q4 2006 and Q1 2007), with a view to adopting them for the e-Infrastructure project in order to make the deployed services available to users registered with Shibboleth-compliant organizations. The WP will draw on the findings of the ASW on Security and Confidentiality to be held in March 2006.

<sup>4</sup> <http://www.athensams.net/news/shibgatewaylaunch.html>

<sup>5</sup> <http://www.sve.man.ac.uk/Research/AtoZ/SHEBANGS>

Deliverable	Description	Due	Effort
D3.2.1	Authentication and authorization services review.	12	3 (+3 Hub resourced)
D3.2.2	Implementation of authentication and authorization service in Web portals.	18	3 (Hub resourced)

**WP3.3 [12 person-months] Portal architecture and support service (CQeSS).** The UK research community is particularly active in using portal technology to support collaborative activities and end-user access to distributed resources in an e-Infrastructure. Some portal development work currently funded by the JISC VRE and e-Science programmes will be used to provide a portal platform for the ESRC e-Infrastructure.

This WP will build on the work being undertaken by CQeSS. CQeSS will work with application developers and data providers to create portal interfaces (using established standards, Java JSR-168 and WSRP) for the tools and services used and produced by NCeSS Hub and Nodes. These interfaces will be designed to meet appropriate usability requirements. Additional services will be provided to host portals as interfaces to the e-Infrastructure at CQeSS and on the NCeSS Hub. The preferred framework is Sakai<sup>6</sup> + uPortal<sup>7</sup>, but other frameworks such as GridSphere, LifeRay or eXo could be used if preferred, or commercial platforms such as Oracle, Sun, BEA or WebSphere.

CQeSS partners (Lancaster/Daresbury) will share expertise and work together to set up the portal framework and establish a repository of re-usable portlet interfaces as part of these ongoing activities. This can include existing open source collaboration and Grid tools developed by the UK community, such as those already in use on the NGS.

Deliverable	Description	Due	Effort
D3.3.1	Portal support/deployment service for NCeSS Hub and Nodes.	9	6
D3.3.2	Usability requirements for portals.	12	2
D3.3.3	A hosted and maintained portal at CQeSS and NCeSS hub.	15	2
D3.3.4	Hosted and maintained portlet repository for tools and services.	18	2

**WP3.4 [6 person-months] Workflow tools (CQeSS).** Kepler and Taverna are examples of a number of widely used open source Grid workflow systems. They include visual editors which provide a simple to use way of composing workflows. The WP will carry out a study of workflow requirements for e-Social Science and use this to select the workflow system best suited to the e-Infrastructure project's needs. The construction of some simple, generic workflows will also be undertaken. The WP will build upon work being undertaken within the Cheshire 3 JISC VRE project.

Deliverable	Description	Due	Effort
D3.4.1	Report on requirements for workflow tools.	6	3
D3.4.2	Integrate and deploy selected workflow tool.	12	3

**WP3.5 [12 person-months] A GROWL architecture and Services (CQeSS).** GROWL (Grid Resources On Workstation Library) is a collection of services that hide the complexity of the Grid from user and provide a secure means of accessing the Grid from their favourite desktop application. This WP will develop extensions to GROWL:

#### Client Side Extensions

1. SQL and XML (OGSA-DAI<sup>8</sup>) plug-in for Stata, SPSS, SAS, R, S and Firefox.
2. Storage Resource Broker (SRB<sup>9</sup>, Grid FTP) plug-in for Stata, SPSS, SAS, R, S and Firefox.
3. SABRE (parallel statistical computing Software) plug-in for Stata, SPSS, SAS, R, S and Firefox (to be funded by CQeSS).

#### Server Side Extensions

<sup>6</sup> <http://sakaiproject.org/>

<sup>7</sup> <http://www.uportal.org/>

<sup>8</sup> <http://www.ogsadai.org.uk/>

<sup>9</sup> [http://en.wikipedia.org/wiki/Storage\\_resource\\_broker](http://en.wikipedia.org/wiki/Storage_resource_broker)

1. Sabre on the NGS (Statistical Computing Service, to be funded by CQeSS).
2. SQL tool for Oracle 9i on the NGS (Data Base Management service on the NGS).
3. SRB on the NGS (Data GRID and Data Archive on the NGS).

Deliverable	Description	Due	Effort
D3.5.1	Requirements gathering, use cases.	3	1
D3.5.2	GROWL architecture specification.	6	2
D3.5.3	A hosted and maintained GROWL service.	9	2
D3.5.4	Portal interface to GROWL server side (SRB, SQL, SABRE) tools.	12	1
D3.5.5	Wrapping of additional applications for the GROWL server side, including remote matrix solvers such as HSL and Net(Grid)Solve.	18	6

**WP3.6 [3 person-months] Technical standards definition (Hub, CQeSS).** This WP will ensure compatibility of technical standards of the e-Infrastructure architecture and its components. It will also establish alignment with the JISC roadmap for a comprehensive Service Oriented Architecture and Framework for Education and Research. R.J. Allan (CQeSS) is on the Joint Framework Working Group and will provide liaison between this and the e-Infrastructure project.

Deliverable	Description	Due	Effort
D3.6.1	Technical standards review.	Continuous	3 (NCeSS resourced)

**WP4 [24 person-months] Deployment of tools and services (Nodes and small grant projects).** Guided by **WP1.2**, a number of tools and services will be selected for deployment over the course of the project.

**WP4.1 [12 person-months] Simulation tool (Moses, PolicyGrid, GeoVue).** This WP will provide an integrated end to end simulation workflow (from data inputs, simulation runs, outputs and analysis) which connects simulations to other Grid resources and services, which could include Grid-enabled census data and/or maps, statistical analysis via GROWL and visualisation (see Figure 2). In this step, we would seek to build on progress from the Hydra PDP in which a simulation workflow was composed from a variety of component services; and from FearlusG in which back-end statistical analysis was incorporated although not Grid-enabled.

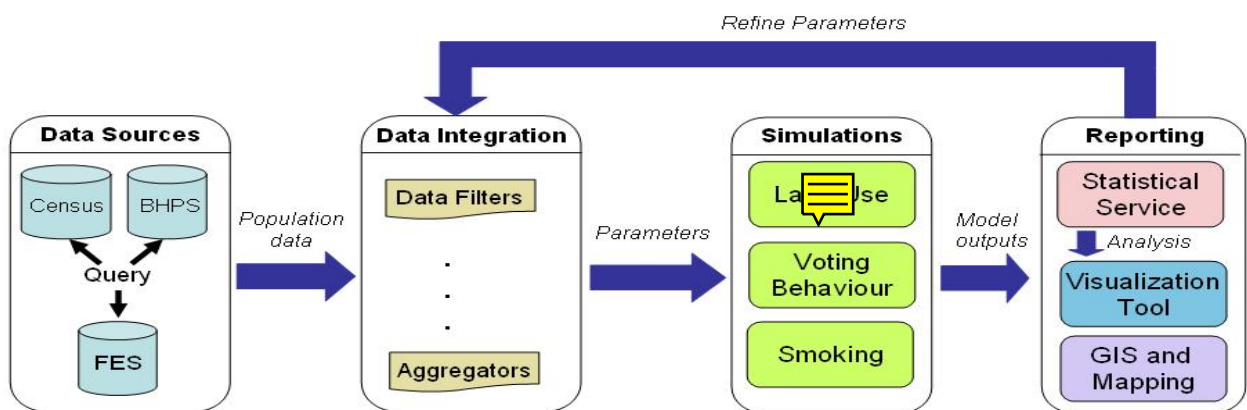


Figure 2: Example Simulation workflows.

As recommended by the Agenda Setting Workshop on e-Infrastructures for Social Simulation, activities within this WP will lead to development of a portal for the simulation community, plus associated archive.

Deliverable	Description	Due	Effort
D4.1.1	Grid-enabling of selected simulation models (Moses, PolicyGrid).	6	2
D4.1.2	Bespoke simulation portal based on WP3.3.	8	2
D4.1.3	Meta-data annotation tools for simulation archive (see WP3.1).	12	4
D4.1.4	Integration of GROWL client library access to remote resources and	15	2



	numerical algorithms.		
D4.1.5	Development of demonstration workflows, e.g., using Kepler (WP3.4).	18	2

**WP4.2 [12 person-months] Demonstrator deployment (Hub, Nodes, small grants).** This WP will select and deploy demonstrators from Nodes (e.g, GeoVue Pollution and Virtual Cities 3D visualisation tools), small grants, and e-Social Science PDPs (see WP 1.2).

Deliverable	Description	Due	Effort
D4.2.1	Selection of demonstrators	9, 18	2
D4.2.2	Portal integration	12, 21	8
D4.2.3	Usability testing and refinement	14, 24	2

**WP5 [4 person-months]: Feasibility study of a Grid-hosted ‘virtual safe setting’ (Hub, OeSS).** This WP will undertake a feasibility study for a pilot virtual safe setting using Grid technologies. Material gathered in this WP will be incorporated into the report under WP7.

**WP5.1 [2 person-month] Confidentiality (OeSS).** We will examine perceptions in the e-Social Science community about issues centering of confidentiality and anonymity. A scoping study will examine perceptions of risk in accessing confidential information via computer networks. Interviews will be conducted with e-Social Science users as well as tool developers. We will also assess whether there are lessons from UK e-Science and from e-research in health that can be extended to e-Social Science.

**WP5.2 [2 person-months]: Security (Hub).** We will review of security frameworks and tools and produce report. It will consult with the National e-Science Centre on Grid security standards, consider the relevance of work by e.g., Elliot et al.<sup>10</sup> and others on technologies for minimizing disclosure risks in the use of confidential data in Grid environments.

Deliverable	Description	Due	Effort
D5.1	Report on perceptions of confidentiality risks.	12	2
D5.2	Report on Grid security frameworks.	12	2 (Hub resourced)

**WP6 [6 person-months] Documentation, user training and support (Hub, Nodes).** This WP will provide training resources to people who want to Grid-enable their data, use tools, services, etc. through online documentation and tutorials, and training events. They will be delivered by appropriate partners in each of the main WPs.

Deliverable	Description	Due	Effort
D6.1	Online training materials.	Continuous	4
D6.2	Training events.	Continuous	2

**WP7 [6 person-months] Evaluation, future development and sustainability (OeSS).** This WP will provide an ongoing evaluation capability to the project, helping to shape recursively its design and conduct. It will provide an analysis of the human, financial and technical resources required to sustain, support and continue to develop e-Infrastructure for the social sciences. The WP will also establish the licensing policy under which the tools developed will be made available. It is critical that licensing policy should both facilitate community engagement and the long-term sustainability of the e-Infrastructure.

**WP7.1 [3 person-months] Ongoing evaluation (Hub, OeSS).** This WP will identify the obstacles to – and successes in – the unfolding programme of work described in the WPs above, through a continuous evaluation of activities being undertaken against objectives, milestones and deliverables. The output will be a series of formative reports to each research team and to the Hub.

**WP7.2 [3 person-months] Sustainability (Hub, OeSS).** All components of e-Infrastructure, once developed, need to be supported if they are to remain of use to the social science research community.

<sup>10</sup> PDP: An investigation of disclosure risks posed by the Grid. Small Grant: The Use of Grid Computing to Facilitate Disclosure Risk Assessment.

Documentation, access, user-support, maintenance, bug-fixes, development are all essential within a service environment. This WP will use experience gained in all the other WPs to analyse the resources required to sustain e-Infrastructure. This will be accompanied by reviewing how sustainability is being addressed in other UK e-Science and world wide initiatives (e.g., NSF, BMFB in Germany) in terms of their plans for integrating, coordinating, and ensuring the longer-term use and viability of e-Social Science.

The main deliverable will be a report summarizing the evaluation which will serve as a guide to future policy and practice for sustainable e-Social Science infrastructure.

Deliverable	Description	Due	Effort
D7.1	Reporting to Hub, Nodes.	8,16,22	3
D7.2	Sustainability Report (includes inputs from WP1).	12, 24	2 (+1 Hub resourced)

**WP8 [6 person-months] Project management (Hub, Nodes).** This WP will provide project management, integration of other WPs and coordination with stakeholders (JISC, UK e-Science core programme, data centres). A Project Board (PB) will be created which will meet every three months by Access Grid Node with face-to-face meetings every twelve months. The PB will report to the NCeSS Research and Service Delivery Boards as appropriate. Its membership will consist of the WP leaders who will also be responsible for coordination of their own sub-WPs.

The PB will ensure a clear assignment of responsibilities for the deliverables. Guided by the outputs of WP7.1, an iterative development process will be followed with a series of meetings to specify and prioritise the use cases to be implemented in the next phase, with the involvement of the social science community.

Deliverable	Description	Due	Effort
D8.1	6 monthly reports.	6,12,18,24	3 (Hub resourced)
D8.2	End of Award report.	30	3 (Hub resourced)

## 6. Budget

The project would start on April 1<sup>st</sup>, 2006 and last for a period of 30 months, with WPs 1-7 collectively running over the first 24 months and WP8 running over 30 months.

The total person-months for the project is 132 (5.5 FTEs for 2 years). We request funding for 114 person-months of RA effort (4.75 FTEs for 2 years) appointed on RA1A scale point 10. The balance of FTEs will be met by existing NCeSS Hub and Node funding.

In addition, we request £12K for equipment/computers and software, and £5K for consumables. We also request £62K for travel and subsistence for workshops, user training events and meetings.

	Year 1		Year 2		Year 3	Total
		x 4.75 FTEs		x 4.75 FTEs		
Staff costs	27200	129,200	28016	133,076		262,276
Superannuation	3808	18,088	3910	18,573		36,661
National Insurance	2064	9,804	2120	10,070		19,874
Overheads @ 46%	15213	72,262	15661	74,391		146,653
Salaries total	48285	229,354	49707	236,109		465,463
Equipment/computers		12,000		0	0	12,000
Consumables		2,500		2,500	0	5,000
Travel and subsistence		30,000		30,000	2000	62,000
<b>Total</b>		273,854		268,609	2000	<b>544,463</b>