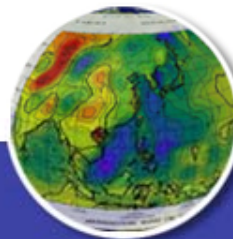




e-Social Science: scaling up social scientific investigations

**Alex Voss, Andy Turner (ESRC National
Centre for e-Social Science)**

**Gabor Terstyanszky, Gabor Szmetanko,
Tamas Kiss (CPC @ Westminster)**



Background

www.euasiagrid.org



- **Social science does not traditionally use advanced ICTs but emergence of new analytical methods is driven by:**
 - Increased availability of data about social phenomena
 - Issues with data management and integration
 - Challenges to analyse social phenomena at scale
 - Challenges to inform practical policy and decision making (e.g., evidence-based policy making)
- **National Centre for e-Social Science (NCeSS) in the UK is investigating ways to respond to these challenges.**
- **EUAsiaGrid is supporting e-Social Science amongst other application domains**

Social Simulation

www.euasiagrid.org



- **MoSeS develops modeling and simulation approaches for social science**
 - First phase research node of NCeSS, now continued through second round node GENeSIS
- **Contemporary demographic modeling of the UK based on UK census data and other datasets**
- **Using agent-based simulation to project population forward in time by 25 years**
- **Simulate the impact of distinct demographic processes such as mortality, fertility, health status, household formation, migration**
- **E.g., to inform policy making – what impact do policy decisions have**

Population Reconstruction

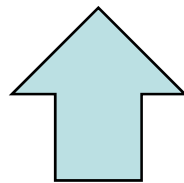
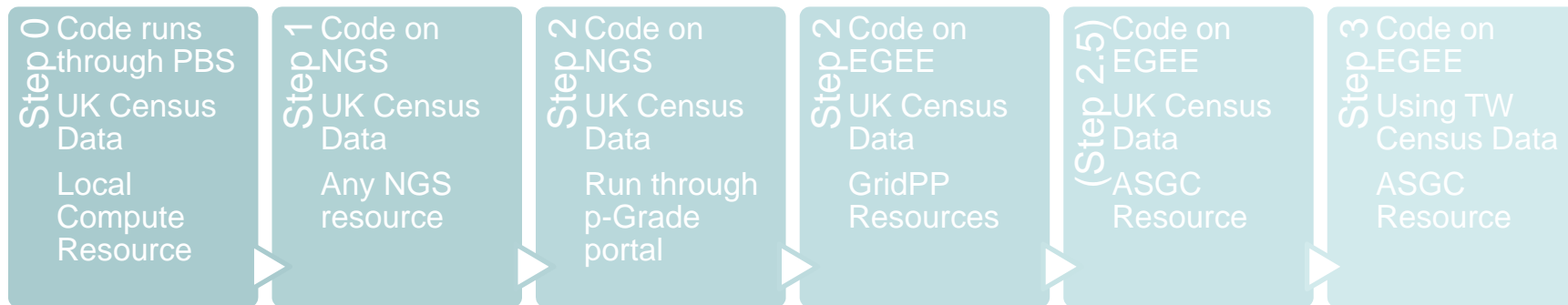
www.euasiagrid.org



- **Generation of a synthetic model of the target population**
 - Based on real census and survey data
 - Works with ‘public release’ versions of census that are restricted, e.g., to 1% of population (anonymisation)
 - Reconstructed model has same attributes as real population and same number of individuals but is still anonymised
 - Can use geographic segmentation to run in parallel
 - Uses genetic algorithms to generate synthetic model
 - Attributes of the model population need to match attributes of the real one, e.g.,
 - overall demographic structure
 - number of students in university areas,
 - number of multi-car households in suburban areas etc.

Application Porting

www.euasiagrid.org



Currently porting population reconstruction code to EGEE, investigating TW data assets and exploring other links, e.g., with healthcare research

Setting up e-Social Science VO

Demo...

www.euasiagrid.org





p-Grade Portal

Workflow Editor - [census_analysis] Mode - View

Workflow Edit Options Help

On 100 25 50 75 100 125 150

```

    graph TD
      Job0[Job0  
GEMLCA] --> Job1[Job1  
MP1]
      Job1 --> Job2[Job2  
SEQ]
  
```

March 2, 2009

System GEMLCA Administration Tool Help

Job Manager

Job list

| Status | [Logs] | [Output] | [Visualization] | [Action] |
|----------|----------|------------|-------------------|-------------------------|
| running | - | N/A | Visualize | All Abort Attach Delete |
| finished | Out - | | | |
| running | -- | | | |
| init | -- | | | |

Done

grid2-portal.cpc.wmin.ac.uk:8080

Start | PGrade Grid portal - Mozilla | blog.doc - Microsoft Word | Document1 - Microsoft Word | Downloads | Workflow Editor - [ce... | 13:37



p-Grade Portal

www.euasiagrid.org



PGrade Grid portal - Mozilla Firefox

File Edit View History Bookmarks Tools Help

https://grid2-portal.cpc.wmin.ac.uk:8080/gridisphere/gridisphere?cid=77&gs_action=doGotoPage

Wireless network we... Getting Started Latest Headlines MoSe5 ESRC e-IP MASS Post to CiteULike [Folder Name] OGC TC 2008-06 Recent Changes in JI...

Microsoft Outlo... INCeS5 Portal Poggio Ducale ... LinkedIn: INSPI... 4th EGEE User ... GridCast Home PGrade Gri... LinkedIn: Reco... GridCast Home Facebook | Inbox

NGS P-GRADE GEMLCA portal

NGS P-GRADE portal University of Westminster

RELEASE 2.4.1G2.0
Information page

Welcome Workflow Certificates Settings OGSA-DAI DataBrowser SRB-Browser Information System GEMLCA Administration Tool Help

Workflow Manager Storage Upload

Workflow Manager

Back Refresh

| Workflow | Job | Gridname | Hostname | Status | [Logs] | [Output] | [Visualization] | [Action] |
|-----------------|------|----------|-----------------|----------|----------|------------|-------------------|---------------------|
| census_analysis | | | | running | - | N/A | Visualize All | Abort Attach Delete |
| | Job0 | NGS | ngs.leeds.ac.uk | finished | Out | - | - | |
| | Job1 | NGS | ngs.leeds.ac.uk | running | -- | - | - | |
| | Job2 | NGS | ngs.leeds.ac.uk | init | -- | - | - | |

Message: Job list refreshed.

March 2, 2009

Done grid2-portal.cpc.wmin.ac.uk:8080

Start PGrade Grid portal - ... blog.doc - Microsoft Word Document1 - Microsoft ... Downloads Workflow Editor - [censu...

13:39



Workflow Manager

Tracefile visualization

workflow: census_analysis

Trace View Info

Job1
ngs.leeds.ac.uk

Job0
ngs.leeds.ac.uk

Width: 600
Height: 350
OK

0s 10s 20s 30s 40s 50s 1m0s

Message: Attempt to visualize successful.

Applet szlaki.trace.client.TraceClientApplet started

grid2-portal.cpc.wmin.ac.uk:8080

Start | PGrade Grid po... | blog.doc - Micros... | Document1 - Micr... | Downloads | Workflow Editor -... | demo-screen2,P... | Workflow Editor -... | 13:40

Detailed description: This is a screenshot of a Mozilla Firefox browser window displaying the PGrade Grid portal. The browser's address bar shows the URL "https://grid2-portal.cpc.wmin.ac.uk:8080/gridsphere/gridsphere?cid=77&gs_action=doVisualizeWorkflowTrace". The page content is titled "Workflow Manager" and "Tracefile visualization". It shows a workflow named "census_analysis" with two jobs, Job0 and Job1, both from "ngs.leeds.ac.uk". A Gantt-style chart visualizes the execution of these jobs over time, with a scale from 0s to 1m0s. Job0 runs from approximately 25s to 55s, and Job1 runs from approximately 55s to 60s. A blue line connects the end of Job0 to the start of Job1. A dialog box on the right allows for adjusting the visualization width (600) and height (350). A message at the bottom states "Message: Attempt to visualize successful." The browser's taskbar at the bottom shows several open applications, including "PGrade Grid po...", "blog.doc - Micros...", "Document1 - Micr...", "Downloads", "Workflow Editor -...", "demo-screen2,P...", and "Workflow Editor -...". The system clock shows 13:40.

Experiences

www.euasiagrid.org



- **Integrating existing code into grid environment required some changes to source code**
 - management of input arguments
 - code scalability
 - log management
 - error handling
- **Finding the right input size and parameters for testing to keep execution times low**
- **Making sense of execution failures**
 - lack of ways to debug code in distributed environments

Experiences II

www.euasiagrid.org



- **Step-wise process works well,**
 - ensures we encounter problems piece by piece
 - allows us to comply with data protection / licensing
- **Population reconstruction is resource intensive**
 - may run up against limits on wall clock time
- **Importance of ‘at elbow’ support**
 - but hindered by data protection/licensing issues
- **Licensing means we need to limit execution to UK resources**
- **No e-Social Science VO for EGEE porting (yet)**
 - Needed to get support from other VOs in the meantime

Future Work

www.euasiagrid.org



- **Next steps until code runs in Taiwan with Taiwanese data**
- **Modularising computation so it can exploit multiple NGS nodes or EGEE CEs**
- **Improving data and code staging**
- **Moving from population reconstruction to supporting the simulation process**
- **Integration into ‘science gateway’ for the social sciences and developing a repository for models**

Acknowledgements

www.euasiagrid.org



- **National Centre for e-Social Science**
 - MoSeS Node: Mark Birkin (PI)
 - GENeSIS Node: Mike Batty (PI)
 - NCeSS Hub: Peter Halfpenny and Rob Procter
- **EUAsiaGrid Consortium**
 - Marco Paganoni (Project Director)
- **CPC at Westminster University**
 - Gabor Szmétankó
 - Gabor Terstyánszky
 - Tamas Kiss
- **GridPP**
 - Jens Jensen and Jeremy Coles
- **National Grid Service**
 - Jason Lander and Shiv Kaushal (Leeds), Steven Young (Oxford), Mike Jones (Manchester)