Modelling and Simulation for e-Social Science (MOSES): A Research Agenda and Progress Report

Mark Birkin

University of Leeds

IBG, September 2005

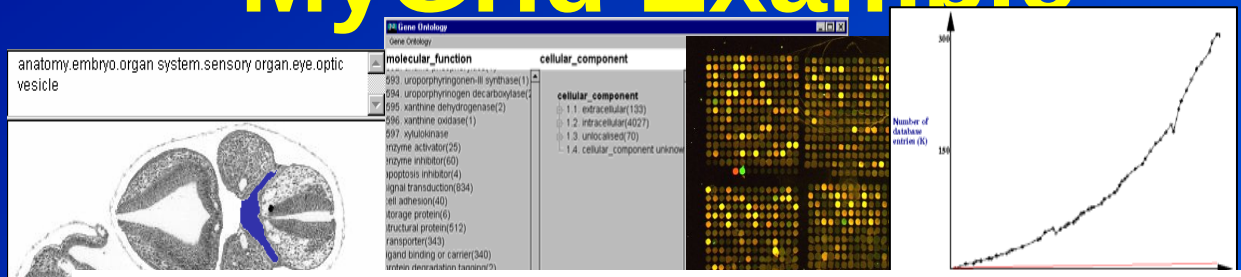
MoSeS

- What is e-Social Science?
- MoSeS – Rationale
- MoSeS – Objectives
- Reflections on Urban Simulation Modelling and SDSS

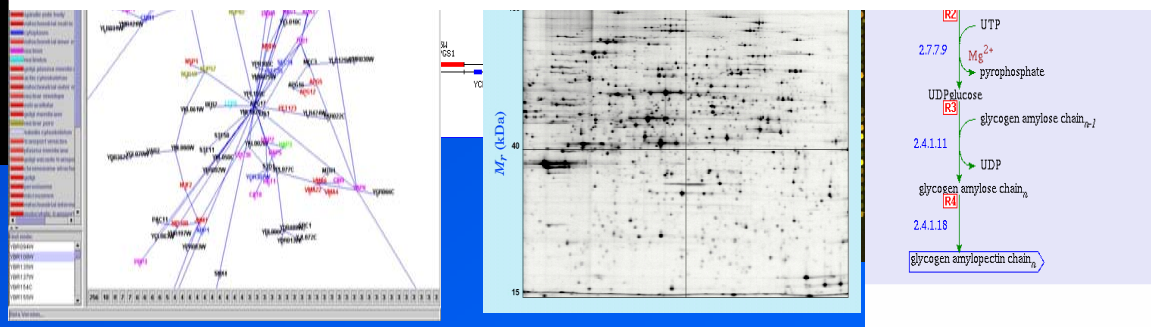
What is e-Social Science

- E-Science
  - Application of Grid Computing to scientific problems
    - Data Grids
    - HPC
  - Enhanced capability for evidence-based science: visualisation, simulation, data mining …
- E-Social Science
  - Application of e-science concepts to social science problem domains

What is e-(Social) Science? MyGrid Example



Motivation: In silico experiments necessitate the virtual organization of people, data, tools and machines. The scientific process also necessitates an awareness of the experience base, both of personal data as well as the wider context of work. The management of all these data and the co-ordination of resources to manage such virtual organizations and the data surrounding them needs significant computational infra-structure support.



What is e-Social Science?
Data Grid

Storage Resource Broker (SRB)

*Data brokered by SDSC instances of SRB***

	As of 1/9/2002		As of 5/17/2002		As of 9/10/2002		As of 5/02/2003		As of 7/24/2003				
Project Instance	*Data_size (in GB)*	*Count (files)*	*Data_size (in GB)*	*Count (files)*	*Data_size (in GB)*	*Count (files)*	*Data_size (in GB)*	*Count (files)*	*Data_size (in GB)*	*Count (files)*	*Users*	*Comments*	*Funding Agency*
NPACI	2,828.18	807,737	1,972.00	1,083,230	2,214.00	1,131,017	4,480.00	1,818,530	6,050.00	2,317,368	367	NPACI Users	NSF/PACI
Digsky	10,565.00	5,079,883	17,800.00	5,139,249	29,006.00	5,225,347	33,930.00	5,292,161	46,100.00	5,719,025	68	2Mass,DPOSS,NVO	NSF/ITR
DigEmbryo	227.77	16,629	433.00	31,629	604.00	43,071	658.00	43,326	720.00	45,365	23	Visible Embryo	NLM
HyperLter	147.50	1,694	158.00	3,596	158.50	3,602	207.00	4,473	215.00	5,097	27	HyperSpectral Images	NSF/NPACI (ESS)
Hayden	3,917.80	18,112	6,800.00	41,391	6,827.00	42,227	7,078.00	59,399	7,078.00	59,399	142	FlyThrough for Planetarium	AMNH/Hayden
Portal	7.40	443	33.00	5,485	53.83	10,278	880.00	24,521	968.00	27,250	316	Grid Portal	NSF/NPACI
SLAC	434.80	9,905	514.00	77,163	605.60	83,839	1,663.00	236,688	1,790.00	254,974	43	Protein Crystallography	NSF/NPACI (Alpha)
NARA/Collection	0.02	301	7.00	2,455	7.80	30,890	47.00	34,077	52.80	79,195	51	Archival Documents	NARA
NSDL/SIO Exp			19.20	383	28.32	1,031	65.12	7,614	232.00	15,809	23	SIO Explorer Documents	NSF/NSDL
TRA			5.80	92	58.50	2,298	91.07	2,371	90.60	2,385	25	Classroom Videos	NSF/NPACI (EOT)
LDAS			239.00	1,766	424.41	3,157	477.86	9,368	498.00	9,858	60	LDAS	
BIRN					66.40	6,600	87.42	177,612	121.00	237,283	138	Biomedical Informatics	NIH (NCRR)
AfCS			27.00	4,007	49.36	6,985	65.80	11,654	95.30	18,762	20	Cell Signalling Images/Docs	NIH
UCSDLib							1,084.00	138,413	1,084.00	138,415	29	Archival Image Files	UCSD
NSDL/CI							177.20	775,959	278.00	993,886	113	K-12 Curriculum Web-sites	NSF/NSDL
SCEC									12.60	18,660	38	South Cal. Earthquake Ctr.	NSF/ITR
TeraGrid									623.00	36,508	1,978	TeraGrid	NSF
TOTAL	18,128.47	5,934,704	28,008.00	6,390,451	40,103.72	6,590,342	50,991.47	8,636,166	66,008.30	9,979,239	3,461		
	18 TB	6 million	28 TB	6.4 million	40.1 TB	6.59 million	51 TB	8.64 million	**66 TB**	**9.97 million**	**3 thousand**		

** Does not cover data brokered by SRB spaces administered outside SDSC.
Does not cover databases; covers only files stored in file systems and archival storage systems

What is e-Social Science?
HPC Grid

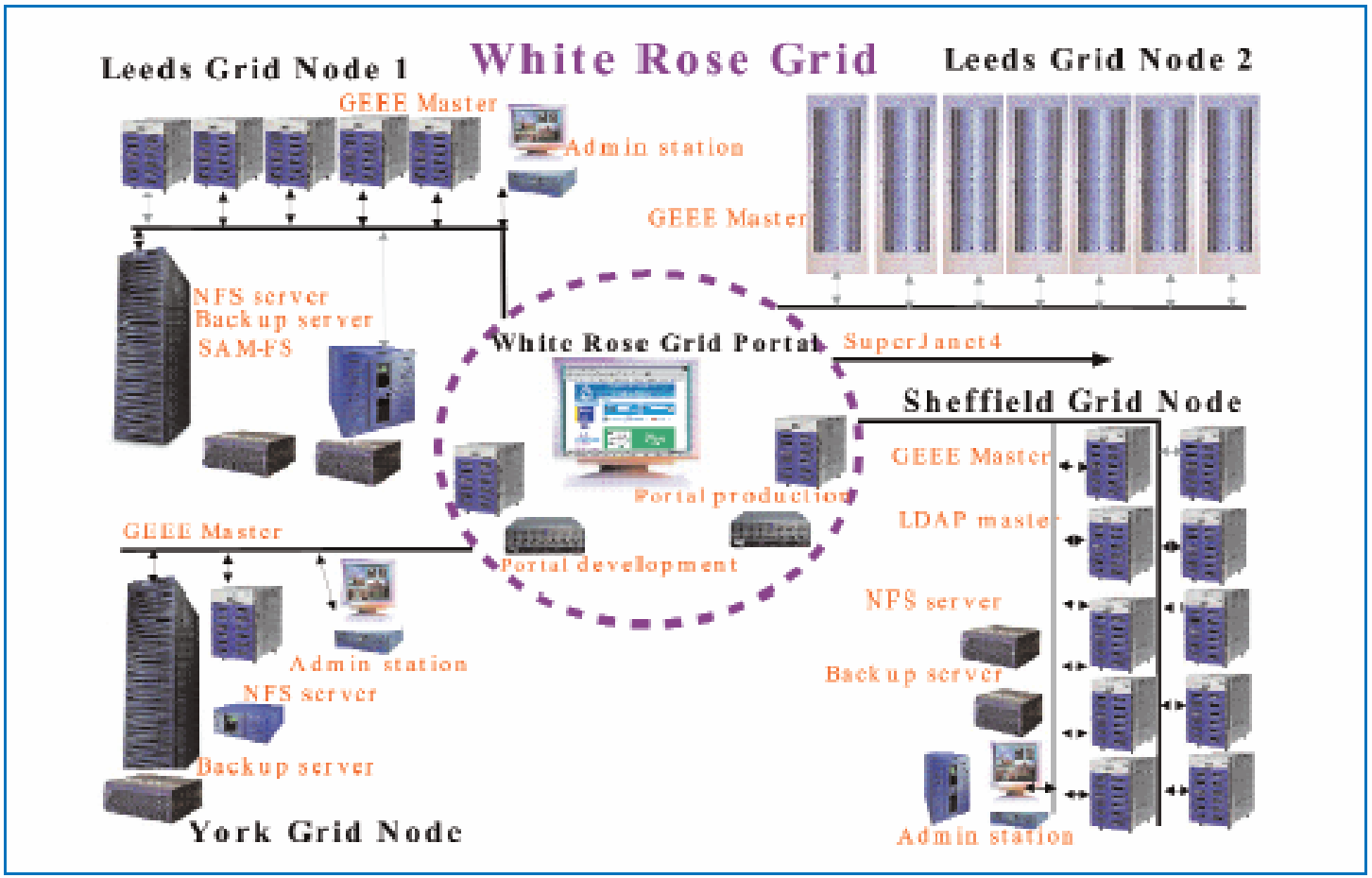
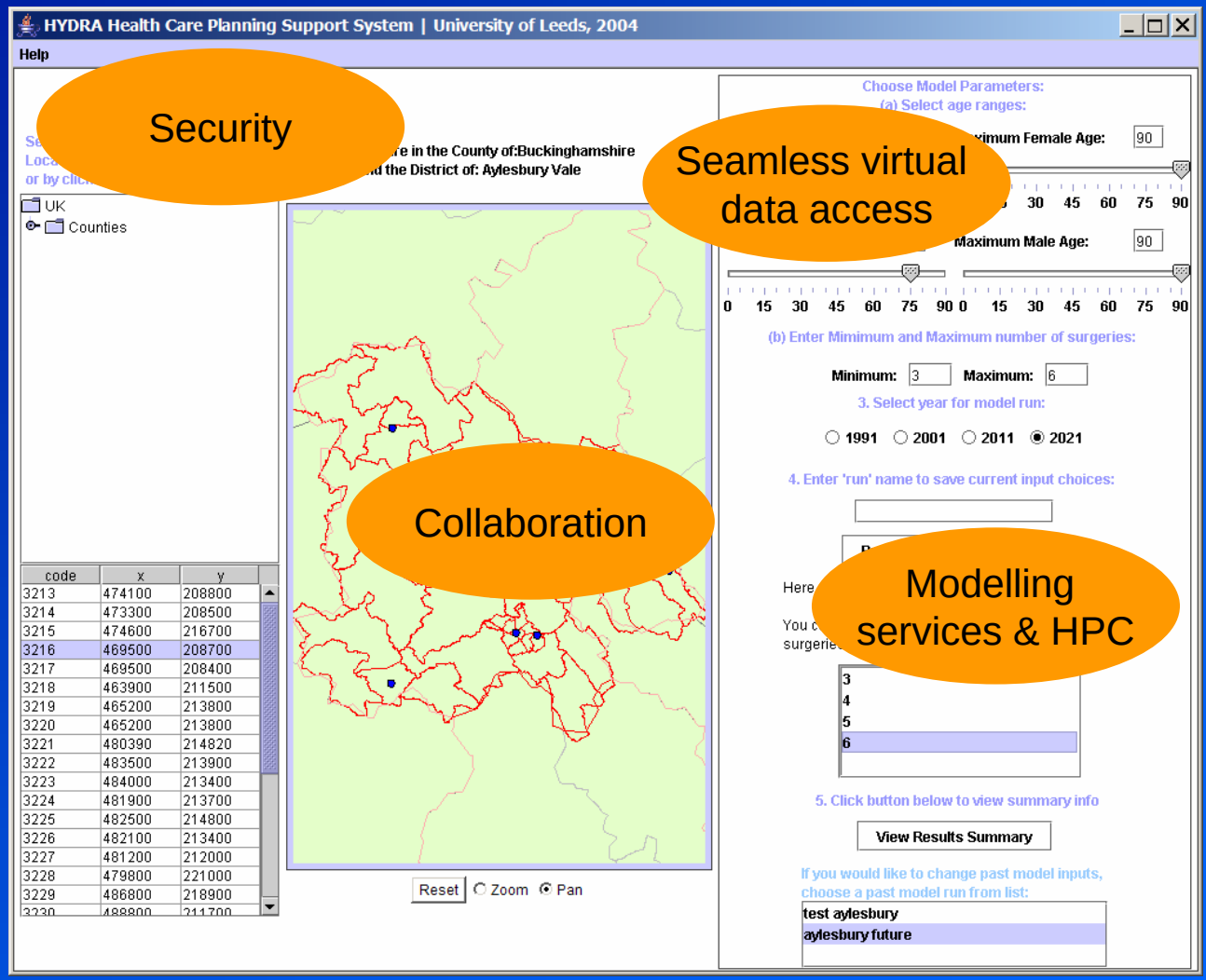


Figure 2: The White Rose Grid Architecture

What is e-Social Science? Hydra Example

- First generation grid-enabled spatial decision support system, using health care scenarios
- Combines virtual database access with spatial mapping, modelling and optimisation tools within a secure open grid services architecture (Globus 3)
- ESRC demonstrator project under the direction of Birkin and Peter Dew

What is e-Social Science?
Hydra Example

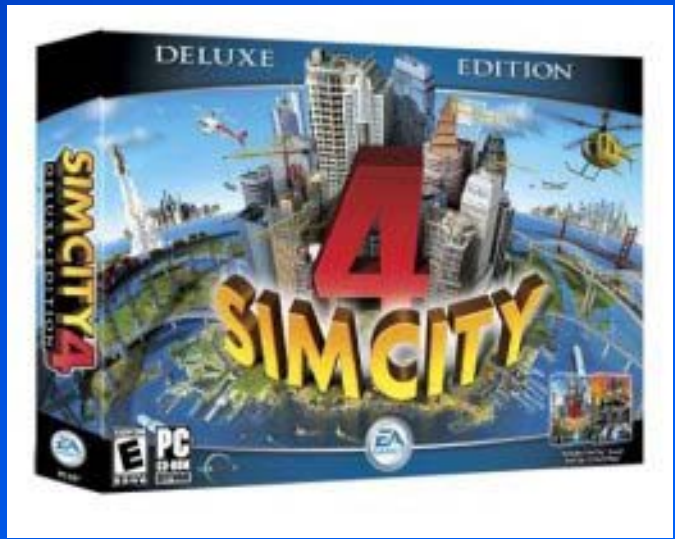


What is e-Social Science?
National Centre for e-Social Science

- An Investigation of Disclosure Issues Posed by the Grid
- Informing Business/Regional Policy: Grid Fusion of Global Data and Local Knowledge (INWA)
- FINGRID: Financial INformation GRID
- SABRE in R: An OGSA Component-Based Approach to Middleware for Statistical Modelling
- Grid-Enabled Micro-Econometric Data Analysis
- **Hydra II Grid Based Spatial Planning Services**
- VIDGRID: Distributed Video Analysis With Grid Technologies
- **Collaborative Analysis of Offenders' Personal and Area-Based Social Exclusion**
- Pilot Semantic Grid Service for Env
- CONVERTGRID
- Genealogies of Knowledge-Develo
  Middleware to Support Fieldwork-E

MoSeS - Rationale

- Suppose that computational power is infinite and spatial data is unlimited
  - What might be achieved under such circumstances?
- Build a SimCity™ ‘for Real’



MoSeS - Rationale

- Urban Simulation Modelling written off by mainstream geography in 1970s
  - But technological conditions now entirely different
  - Imperative to review previous assumptions
- Experience of GMAP in the private sector demonstrates the applied value of spatial analysis and modelling to location planning

Moses - Objectives

- Development of a national demographic model specified for a synthetic population of individuals and households
  - Components include estimation, updating, forecasting and activity modelling
- Application of the model to policy domains including health, business, and leisure

Moses - Methodology

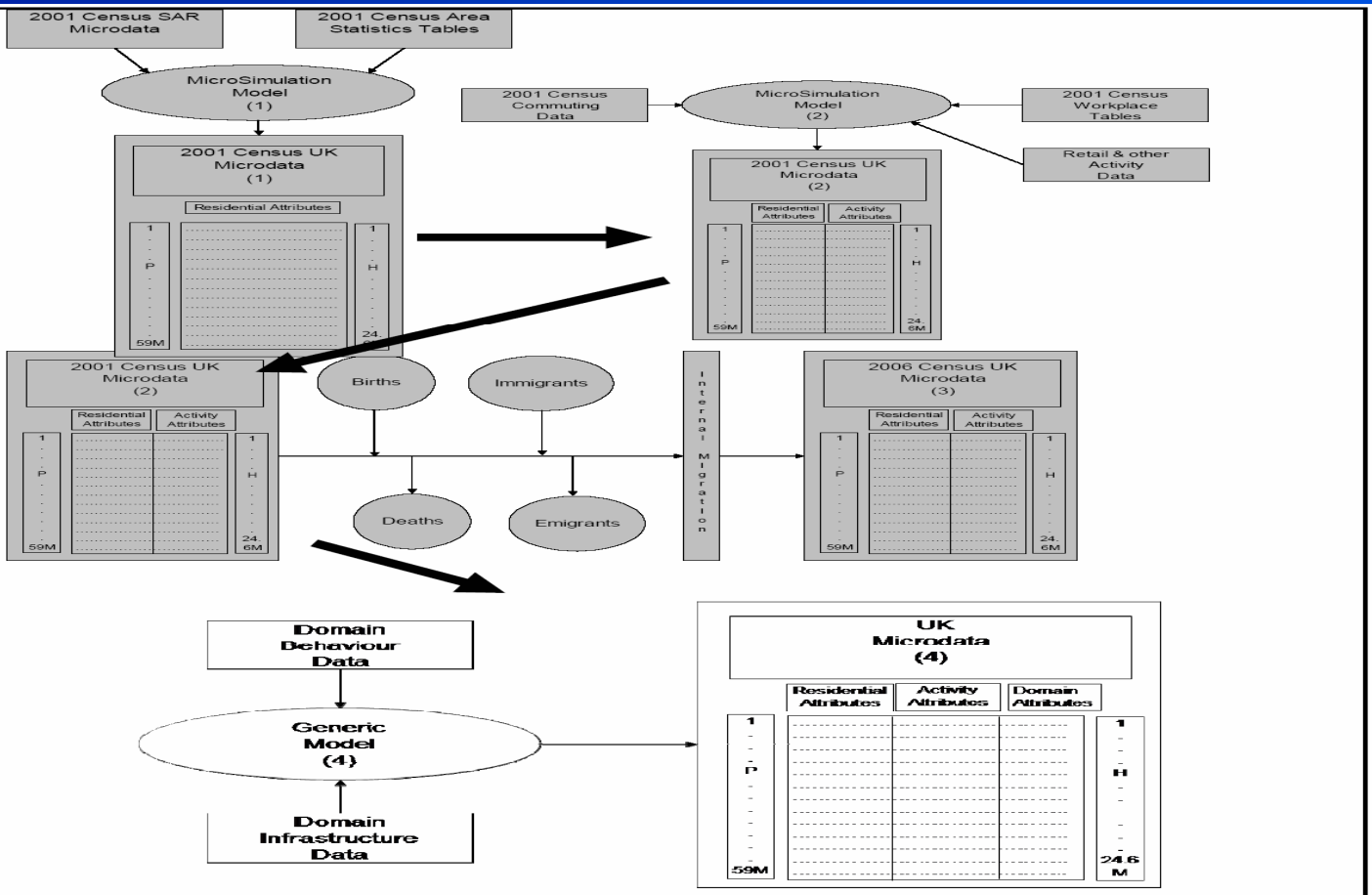


Figure 6.4 Shows how attributes from any social science domain could be added to the model.

Moses - Applications

- Goal is to look at the balance of service provision across both the health and social care sectors
  - Important policy implications due to poor integration between these sectors
  - Increasingly problematic in particular with respect to the very elderly
  - Geographical variation as variations in use will reflect variations in provision: different demographic groups may also demand alternative service mix
  - Possible importance of ‘social networks’ – voluntary services, church, school, health clubs and centres – may have subtle and important influence
  - Problem domain of interest to geography, health economics, political science and social policy
  - Practical importance to Health Care Commission, CSCI, Local Government/ Social Services, Hospital Trusts, Primary Care Trusts …
  - Important dimension of data sharing, confidentiality and security…

Reflections and Ambitions

- Modelling is a rich methodology:
  - Applied value
  - Pedagogic value
  - Interpretive value
  - Scholastic value

Reflections and Ambitions

- Applied value of modelling
  - ‘what if’ modelling; impact analysis
  - forecasting
  - optimisation
  - resource allocation
  - backcasting

Reflections and Ambitions

- Pedagogic value of modelling
  - SDSS concept
  - Provide basis for exploration, learning and understanding of application scenarios
  - Robustness, sensitivity analysis, …
  - Simulations as narrative?

Reflections and Ambitions

- Interpretive value of modelling
  - distribution channels example
  - question and review representation of geographical process or structure
  - revise assumptions and understanding of mechanisms

Reflections and Ambitions

- Scholastic value of modelling
  - agents example
  - city simulations as Big Science/ Grand Challenges
  - substantial intellectual and academic endeavour