

DIGITAL GEOGRAPHY *in a* WEB 2.0 WORLD



WEDNESDAY 20 FEBRUARY 2008

WEDNESDAY 20 FEBRUARY 2008 @ The Barbican

MoSeS:

SimCity for Real

Martin Clarke & Mark Birkin

School of Geography, University of Leeds

<http://www.leeds.ac.uk/geog>

Contents

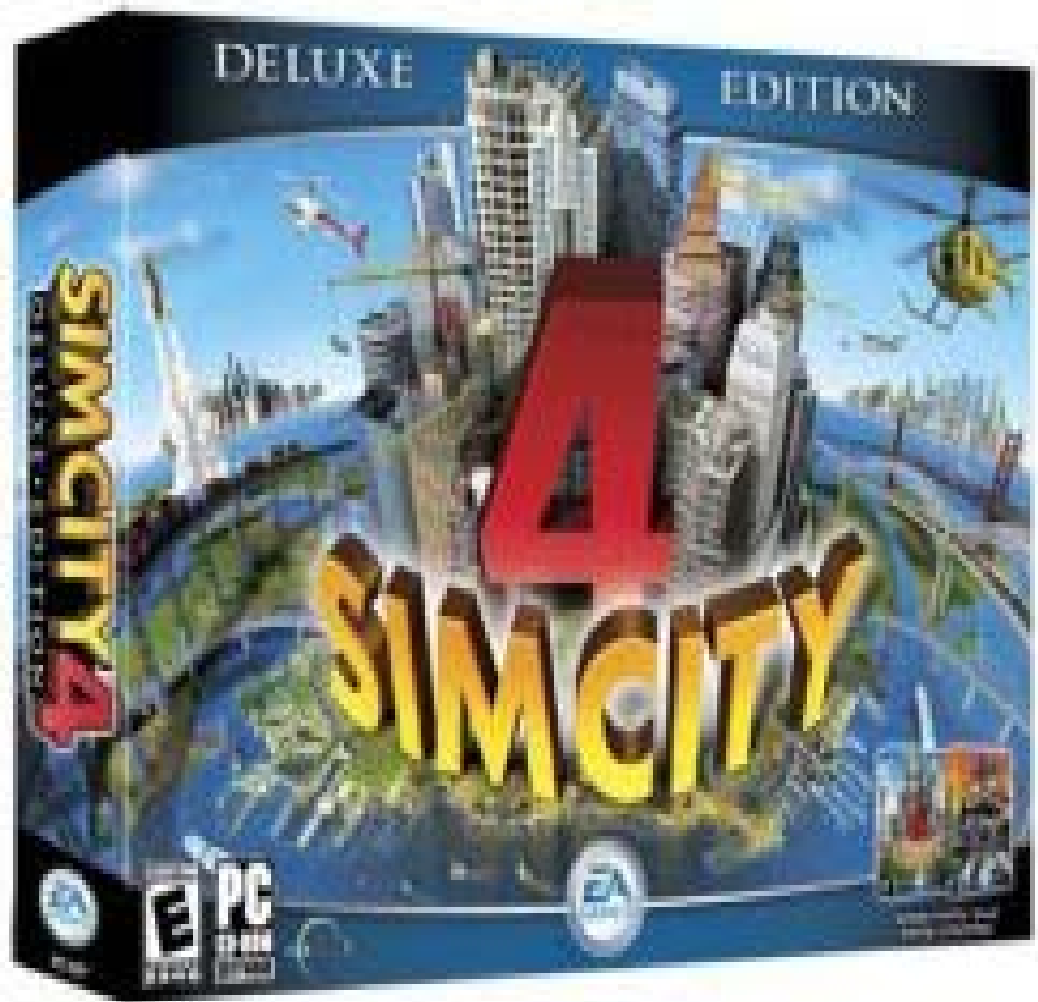
- Background
- Aims
- Generating an Initial Population
- Dynamic Modelling
- Applications
- Architecture
- Next Steps

50 Years of Urban and Regional Modelling

- despite the rich tradition of urban modelling over the last half century there has been relatively limited success in application
 - complex systems
 - lack of good data
 - computational power
 - ability to represent policy issues in a realistic way
- BUT: - this could be beginning to change!

SimCity

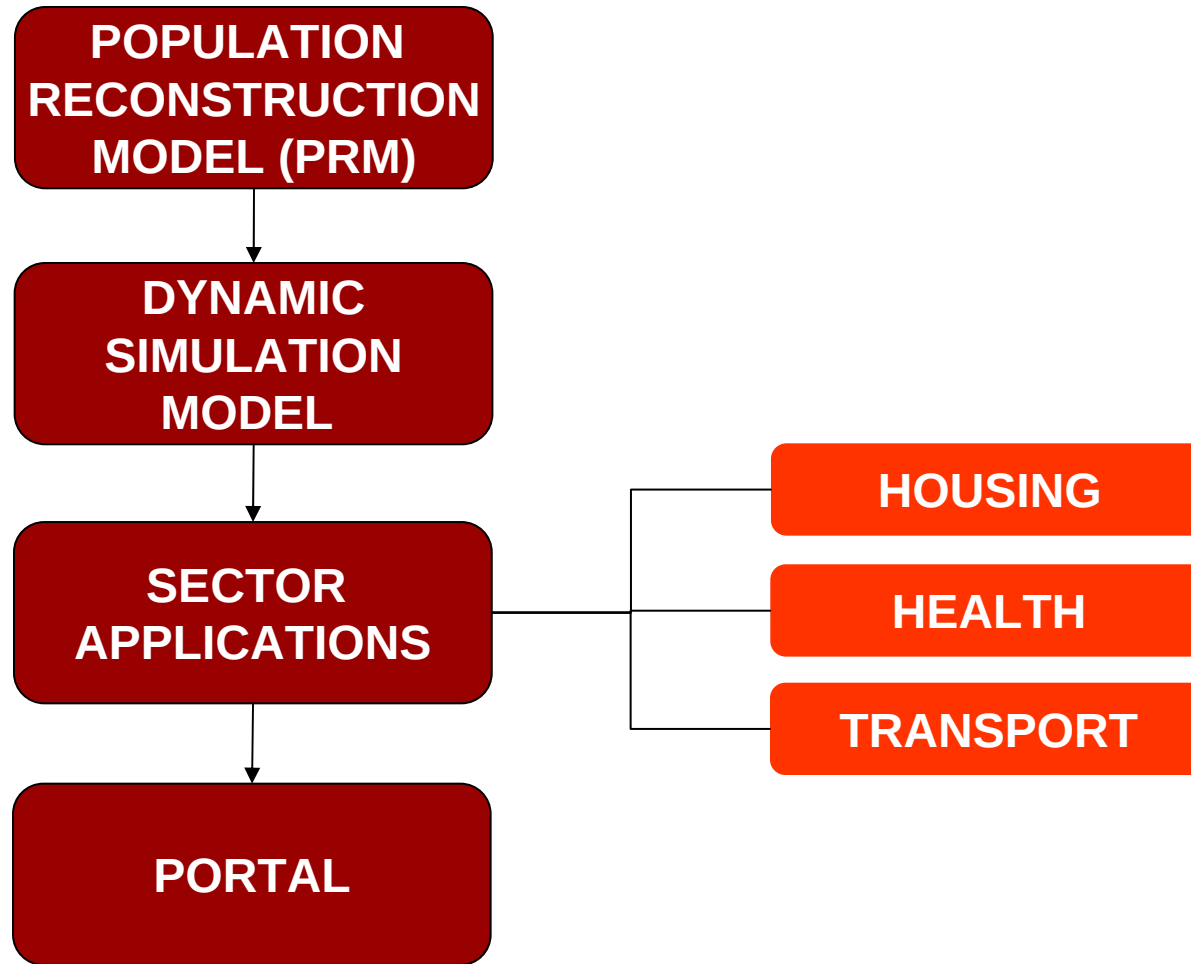
- Popular PC simulation game
- Excellent graphics and user interface
- Plausible logic to urban dynamics
- Available now for 'real' cities
- Can we create a "SimCity for Real"?



MoSeS – Modelling and Simulation for e-Social Science

- Part of the ESRC NCeSS programme
- Challenge of constructing simulations which represent the behaviour of real people in real cities.
- Construct an e-infrastructure for urban simulation
- Deployment of various demographic models of population composition and its change through time
- Applications of the simulation approach to problems in planning health care, housing and transportation

MoSeS: Schematic Overview



Generating an Initial Population

- The Population Reconstruction Model (PRM) constitutes a complete synthetic representation of the UK population, represented as a series of individual people, grouped into households and represented at the finest possible spatial scale – Output Areas (200,000 in England & Wales).
- The base data are drawn from the UK Census Sample of Anonymised Records (SAR) and embellished by other data sets such as BHPS.
- PRM comprises totally ‘realistic’ individuals and household relationships

with no possibility that real named individuals can be identified within the synthetic population.

Generating an Initial Population

- We know the aggregate population and socio-economic characteristics of Output Areas – these act as constraints in the PRM process
- We make use of micro-data from individual and household SARs to populate an Output Area with people, who when looked at in aggregate accurately replace the known cross-tabs
- Use a genetic algorithm to optimise this process – computationally intensive!

Methodology for the PRM

Census tables for zone i

Household SAR

BHPS

Repeat this process for 52 million individuals

DIGITAL GEOGRAPHY in a WEB 2.0 WORLD

NCeSS UCL E.S.R.C. ECONOMIC & SOCIAL RESEARCH COUNCIL UNIVERSITY OF LEEDS SPLINT Spatial Literacy IN Teaching

Methodology for the PRM

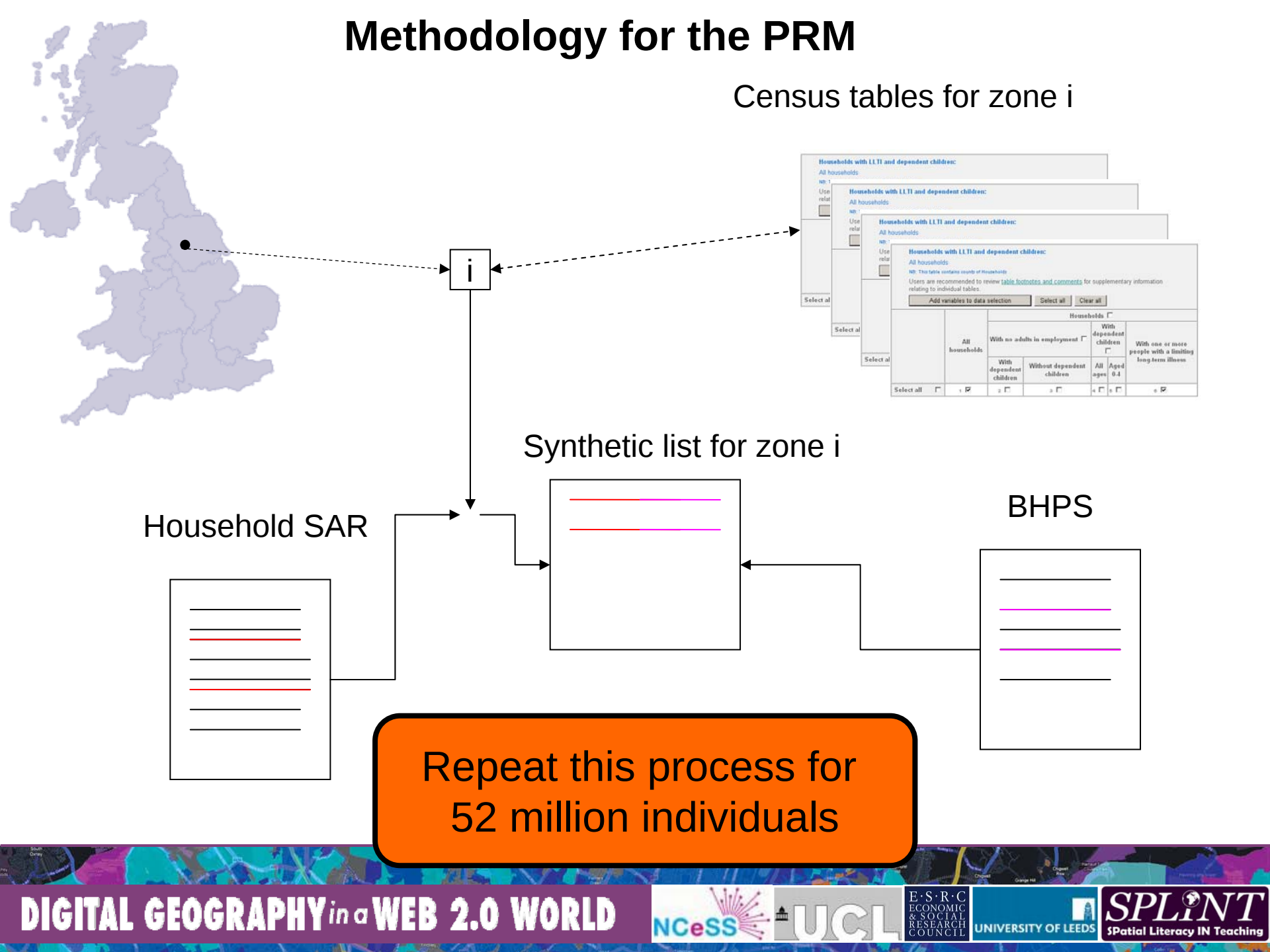
Census tables for zone i

Household SAR

BHPS

Repeat this process for 52 million individuals

Detailed description: The diagram shows a workflow for generating synthetic data. It begins with a map of the United Kingdom on the left. A dashed arrow points from a specific location on the map to a box labeled 'i'. From box 'i', a solid arrow points down to a box labeled 'Synthetic list for zone i'. To the right of box 'i', there is a stack of four overlapping screenshots of census data tables, with the top one titled 'Households with L1 TI and dependent children:'. A dashed arrow points from the top screenshot to box 'i'. Below box 'i', there are three boxes representing data sources: 'Household SAR' on the left, 'BHPS' on the right, and a central 'Synthetic list for zone i' box. Arrows point from 'Household SAR' and 'BHPS' to the central 'Synthetic list for zone i' box. At the bottom, a large orange rounded rectangle contains the text 'Repeat this process for 52 million individuals'. The bottom of the slide features a banner with logos for 'DIGITAL GEOGRAPHY in a WEB 2.0 WORLD', 'NCeSS', 'UCL', 'E.S.R.C ECONOMIC & SOCIAL RESEARCH COUNCIL', 'UNIVERSITY OF LEEDS', and 'SPLiNT Spatial Literacy IN Teaching'.



Methodology for the PRM

Census tables for zone i

Household SAR

BHPS

Repeat this process for 52 million individuals

DIGITAL GEOGRAPHY in a WEB 2.0 WORLD

NCeSS UCL E.S.R.C. ECONOMIC & SOCIAL RESEARCH COUNCIL UNIVERSITY OF LEEDS SPLINT Spatial Literacy IN Teaching

Methodology for the PRM

Census tables for zone i

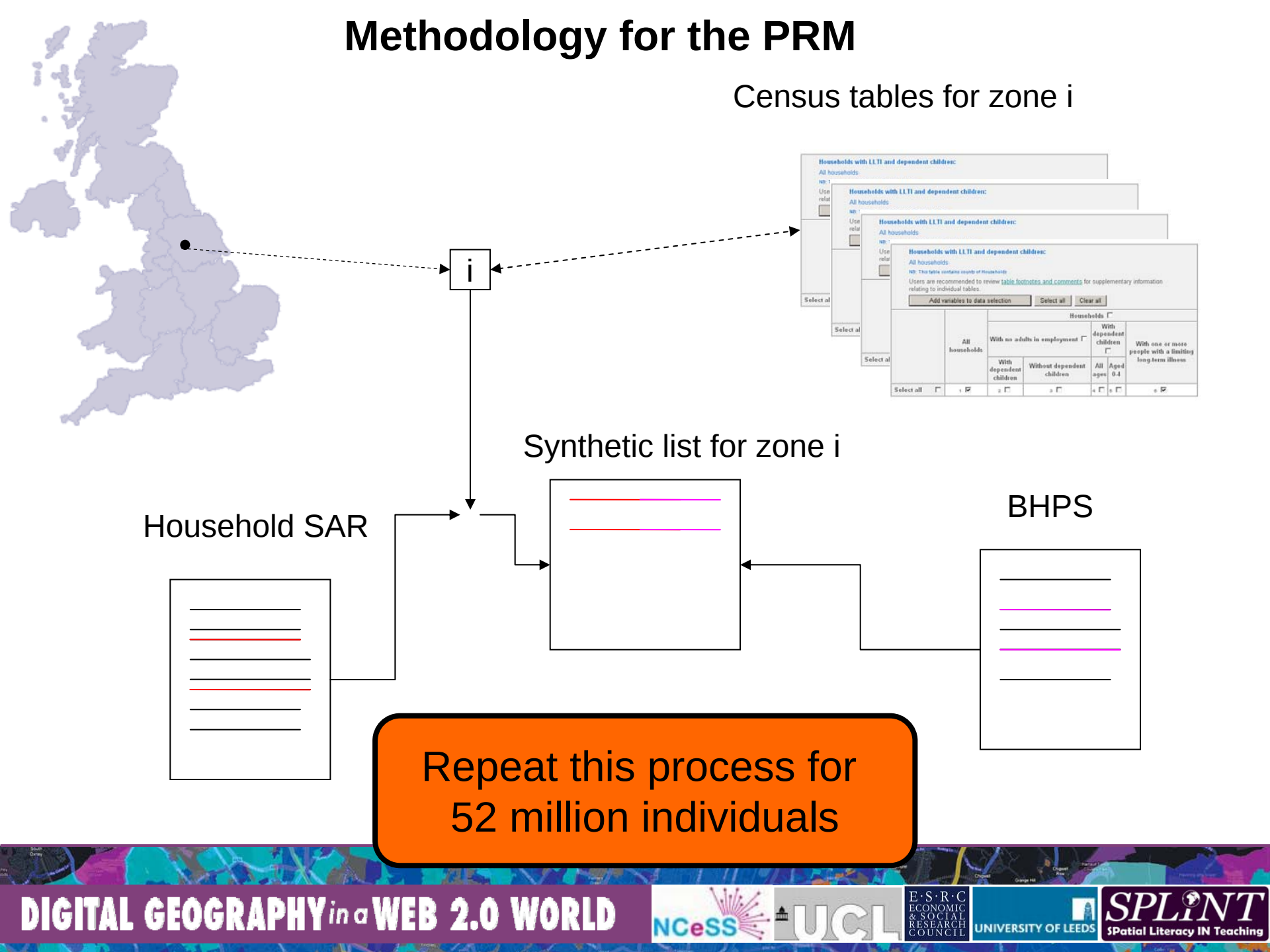
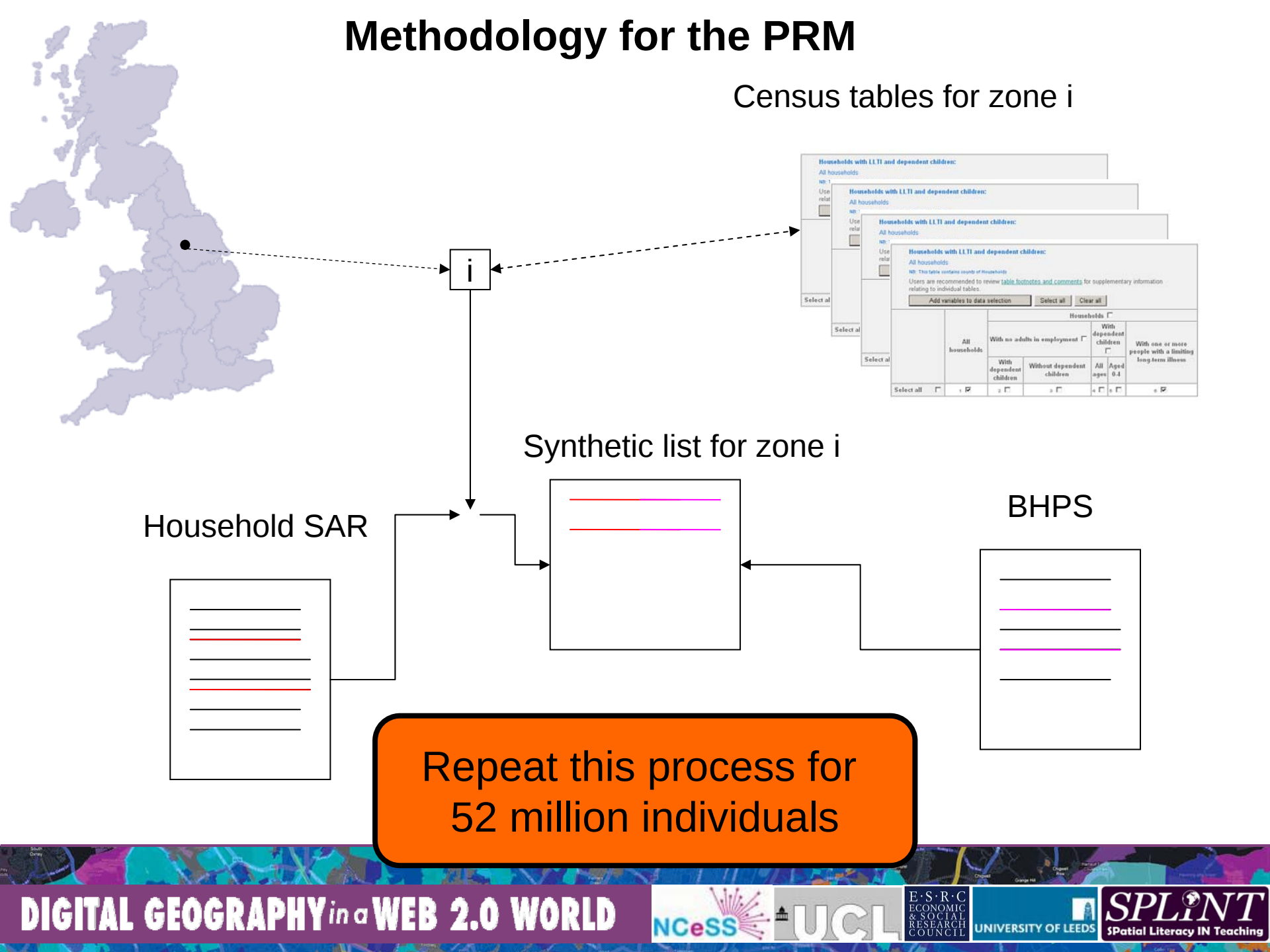
Household SAR

BHPS

Repeat this process for 52 million individuals

DIGITAL GEOGRAPHY in a WEB 2.0 WORLD

NCeSS UCL E.S.R.C. ECONOMIC & SOCIAL RESEARCH COUNCIL UNIVERSITY OF LEEDS SPLINT Spatial Literacy IN Teaching



Methodology for the PRM

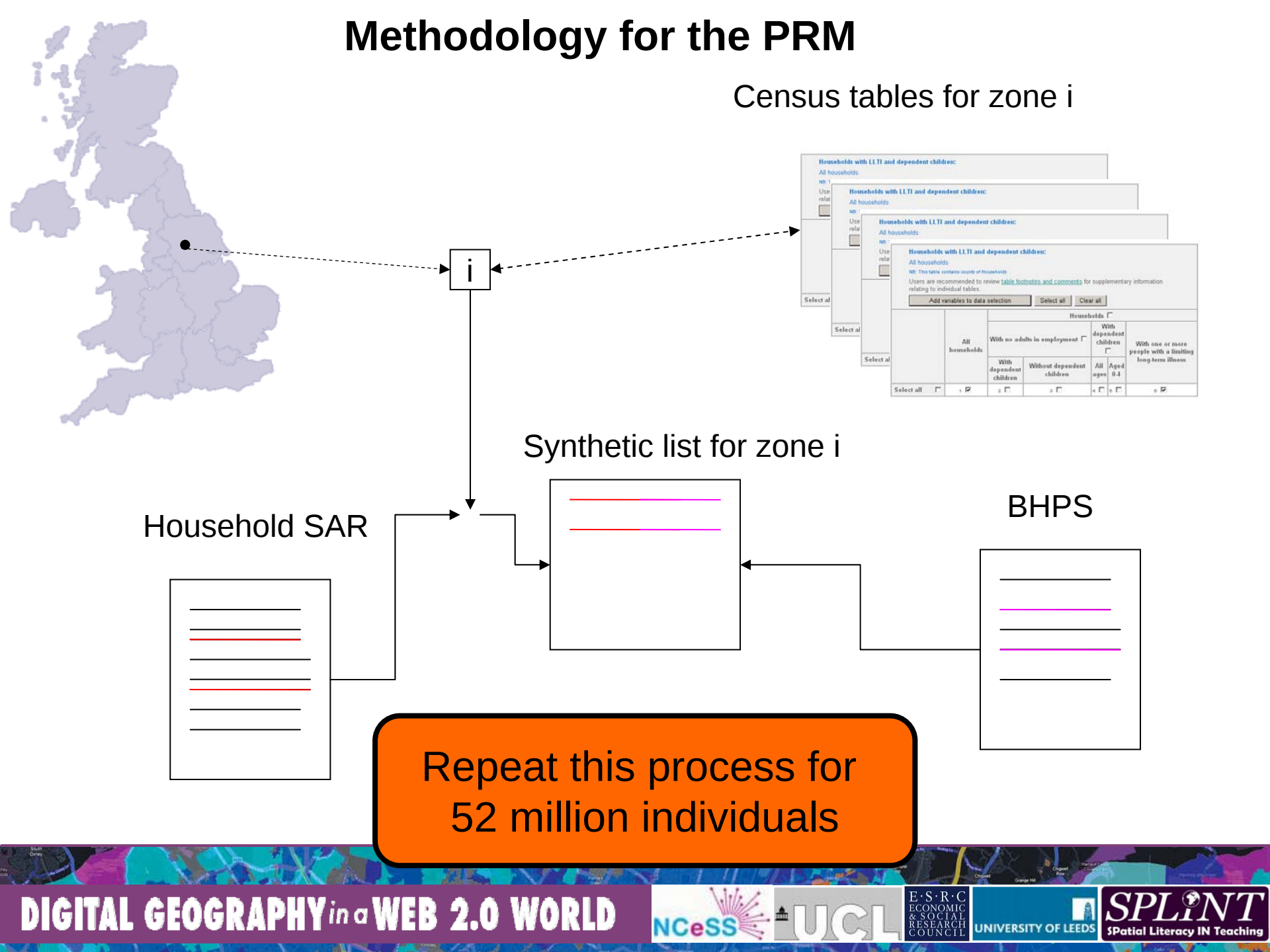
Census tables for zone i

Household SAR

BHPS

Repeat this process for 52 million individuals

Detailed description: The diagram illustrates a methodology for the PRM (Personalized Random Match) process. It begins with a map of the United Kingdom, where a specific geographic zone 'i' is identified. This zone is linked to 'Census tables for zone i', represented by a stack of four tables. These tables feed into a 'Synthetic list for zone i', depicted as a box containing two horizontal lines. The 'Synthetic list' is then used to generate a 'Household SAR' (Spatially Annotated Record) and a 'BHPS' (British Household Panel Survey) dataset, each shown as a box with horizontal lines. The 'Household SAR' and 'BHPS' datasets are then combined to form the final 'Synthetic list for zone i'. The process is repeated for 52 million individuals, as indicated by the orange box at the bottom.



Methodology for the PRM

Census tables for zone i

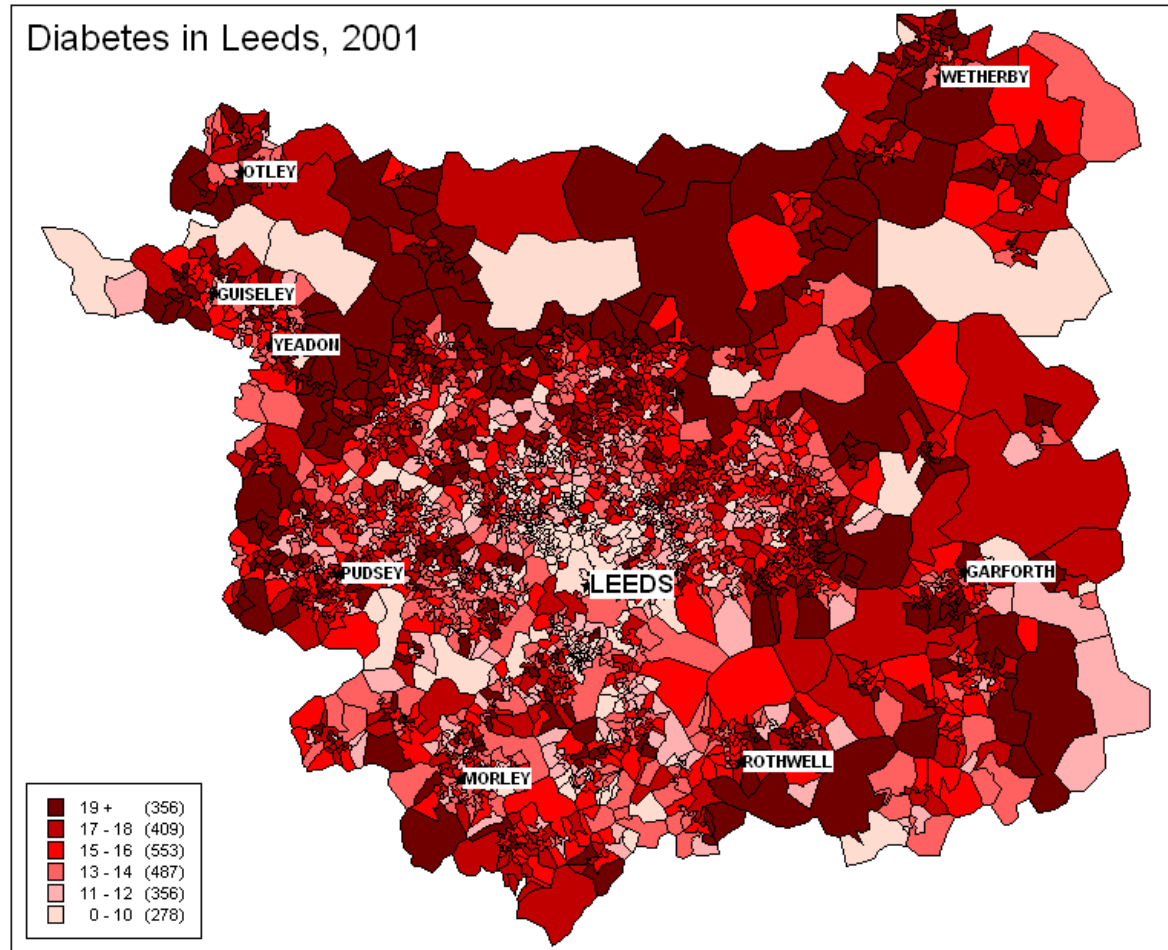
Household SAR

BHPS

Repeat this process for 52 million individuals

Detailed description: The diagram illustrates a methodology for the PRM (Personalized Random Match) process. It begins with a map of the United Kingdom, where a specific geographic zone 'i' is identified. This zone is linked to 'Census tables for zone i', represented by a stack of four tables. These tables feed into a 'Synthetic list for zone i', depicted as a box containing two horizontal lines. The 'Synthetic list' is then used to generate a 'Household SAR' (Spatially Annotated Record) and a 'BHPS' (British Household Panel Survey) dataset, each shown as a box with horizontal lines. The 'Household SAR' and 'BHPS' datasets are then combined to form the final 'Synthetic list for zone i'. The process is repeated for 52 million individuals, as indicated by the orange box at the bottom.

Generating an Initial Population



- By combining different microdata sets we can produce plausible estimates of distribution that are not available from conventional sources

Dynamic Modelling

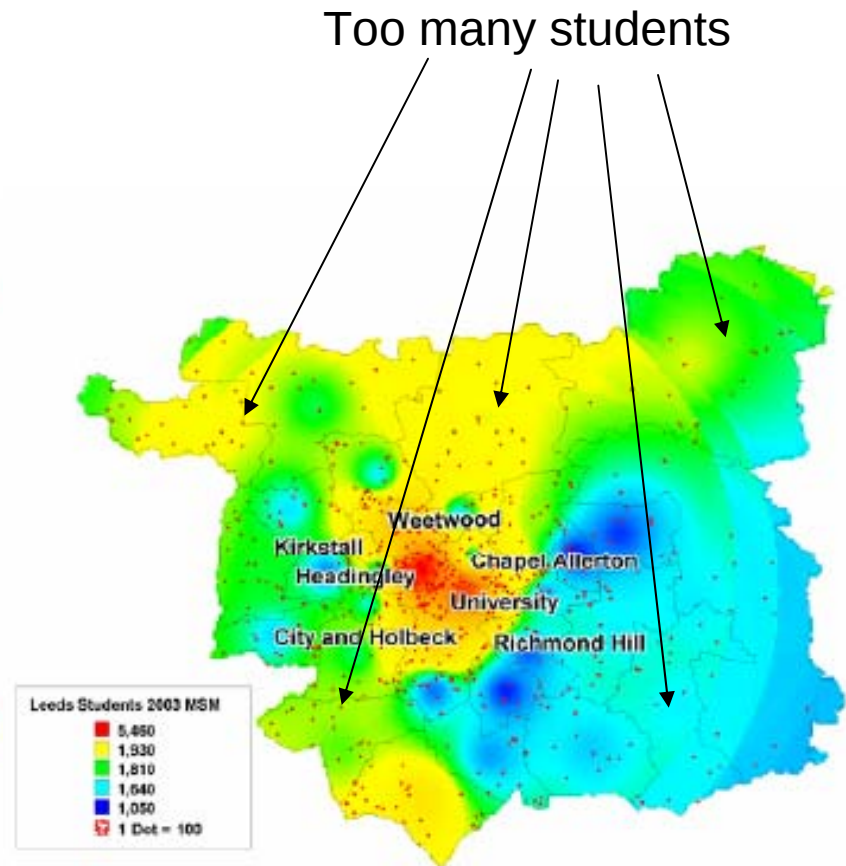
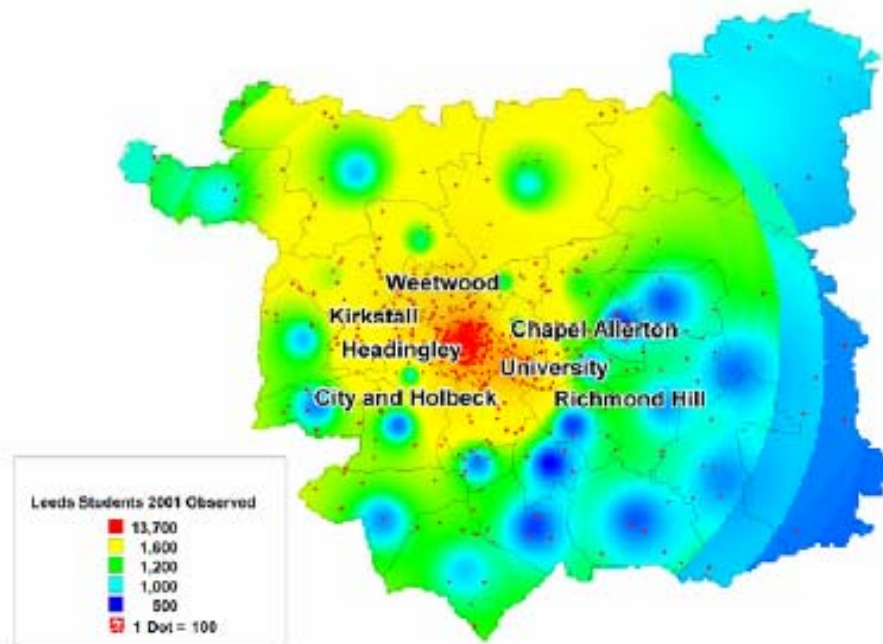
The dynamic model projects the synthetic population (PRM) into the future.

This is achieved through the simulation of individual demographic processes, including births, deaths, migration, marriage, separation, and changes in health status.

Dynamic model has been implemented up to 2031

Computationally intensive

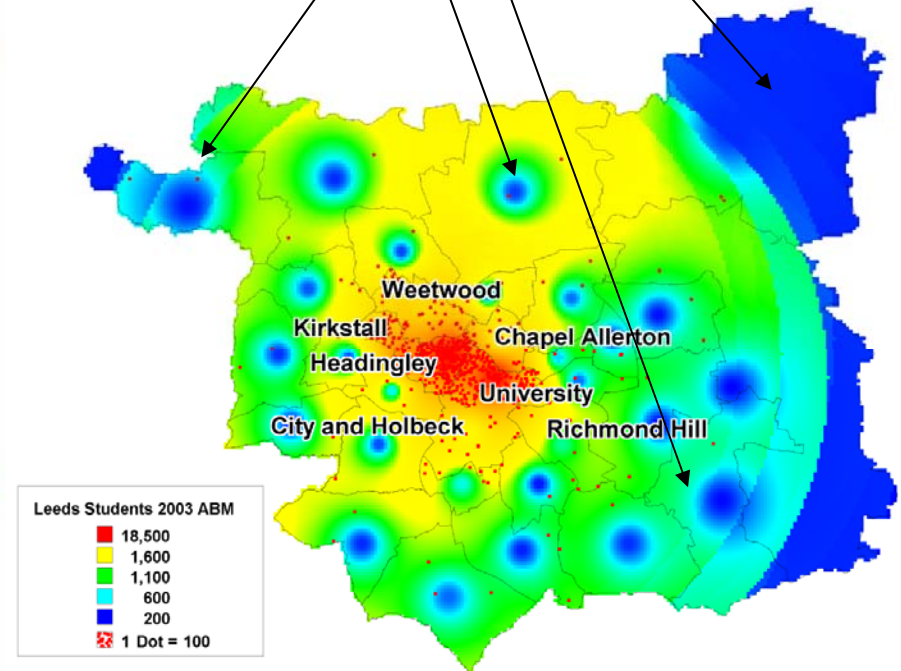
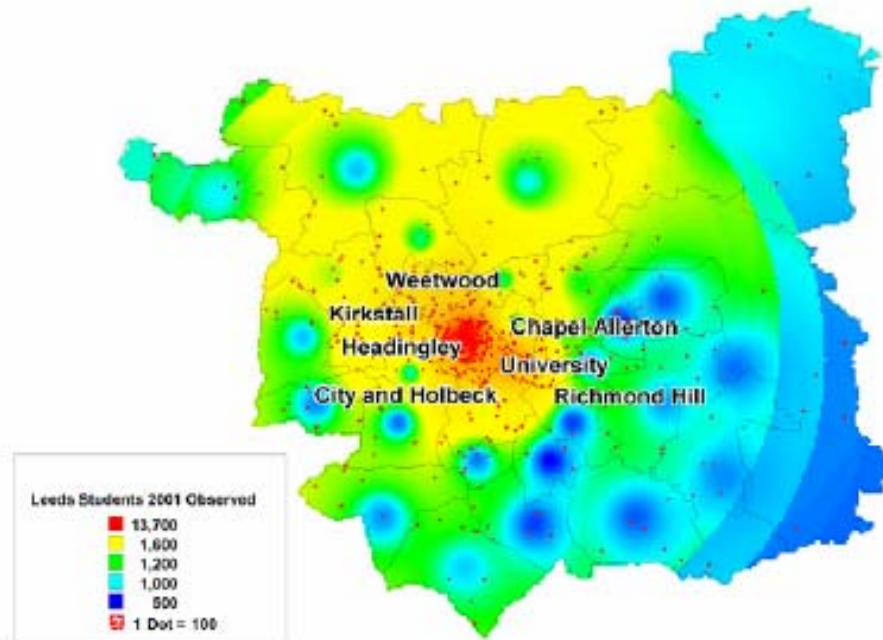
Dynamic Modelling



Location of student populations in a conventional model – an acute example of the difficulties in simulating intra-urban migration patterns

Dynamic Modelling

Correct number/ distribution of students



Location of student populations – use of agent-style rules (no marriage or family relationships; three year degree courses; preference for Halls of Residence...) allows much more realistic behaviours

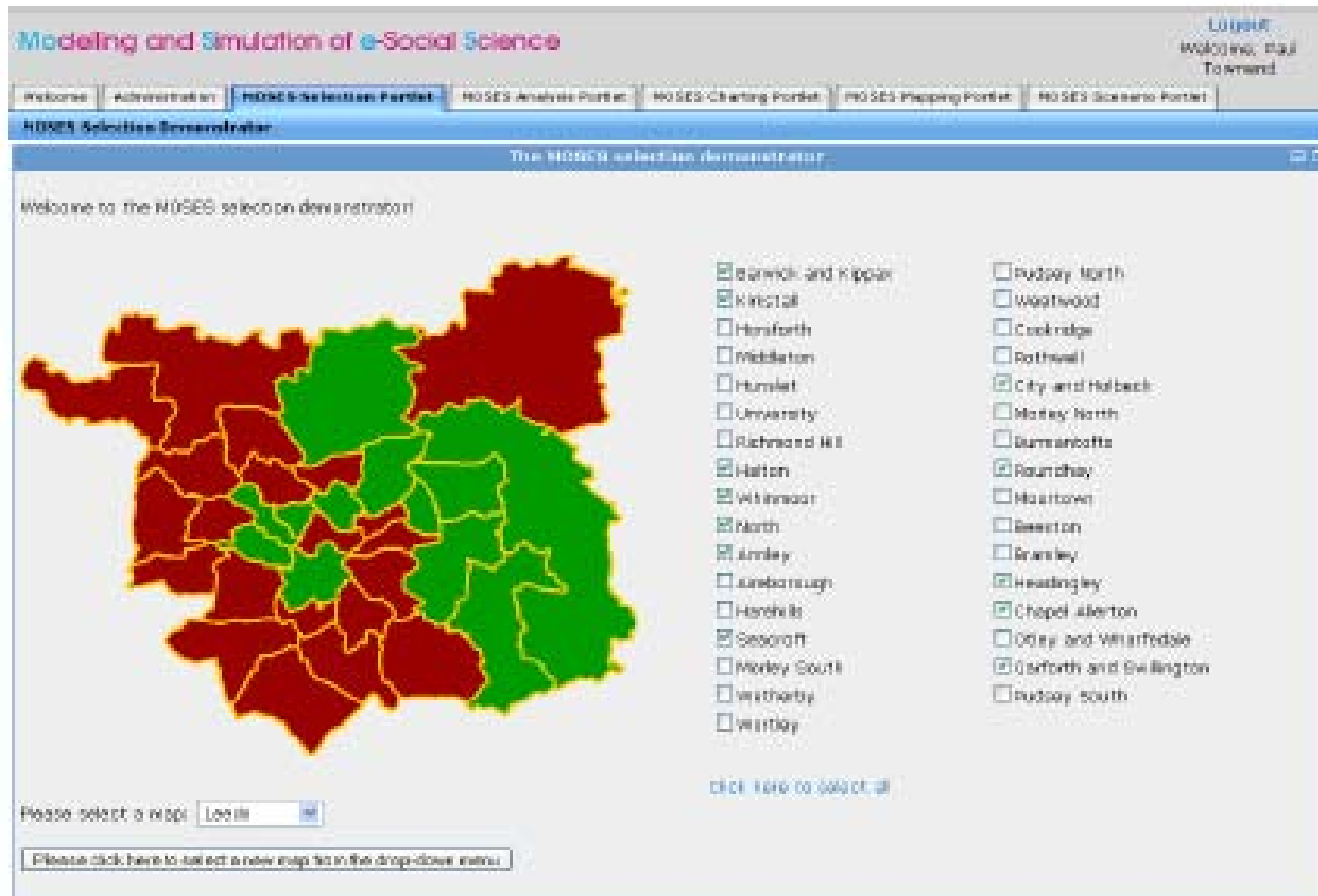
Applications

Moses is conceived as not just an academic and intellectual exercise, but as a practical tool for policy users.

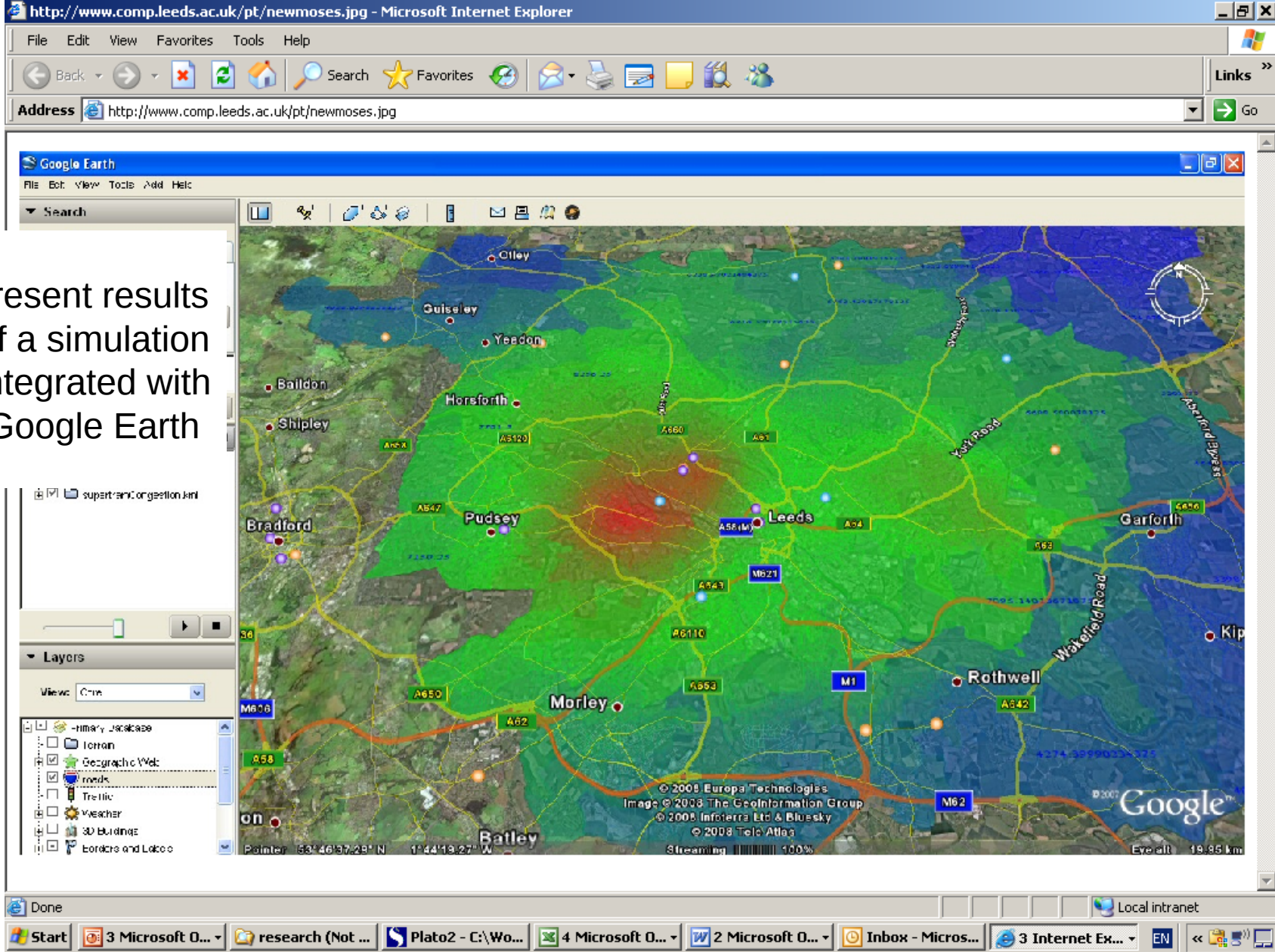
Moses is building local relationships with interested users in the Leeds Teaching Hospitals Trust, Leeds City Council Social Services, and the Leeds Area Regeneration Partnership.

These ideas are also being shared with a broader nationwide community and with a global audience of planners and policy-makers.

Applications



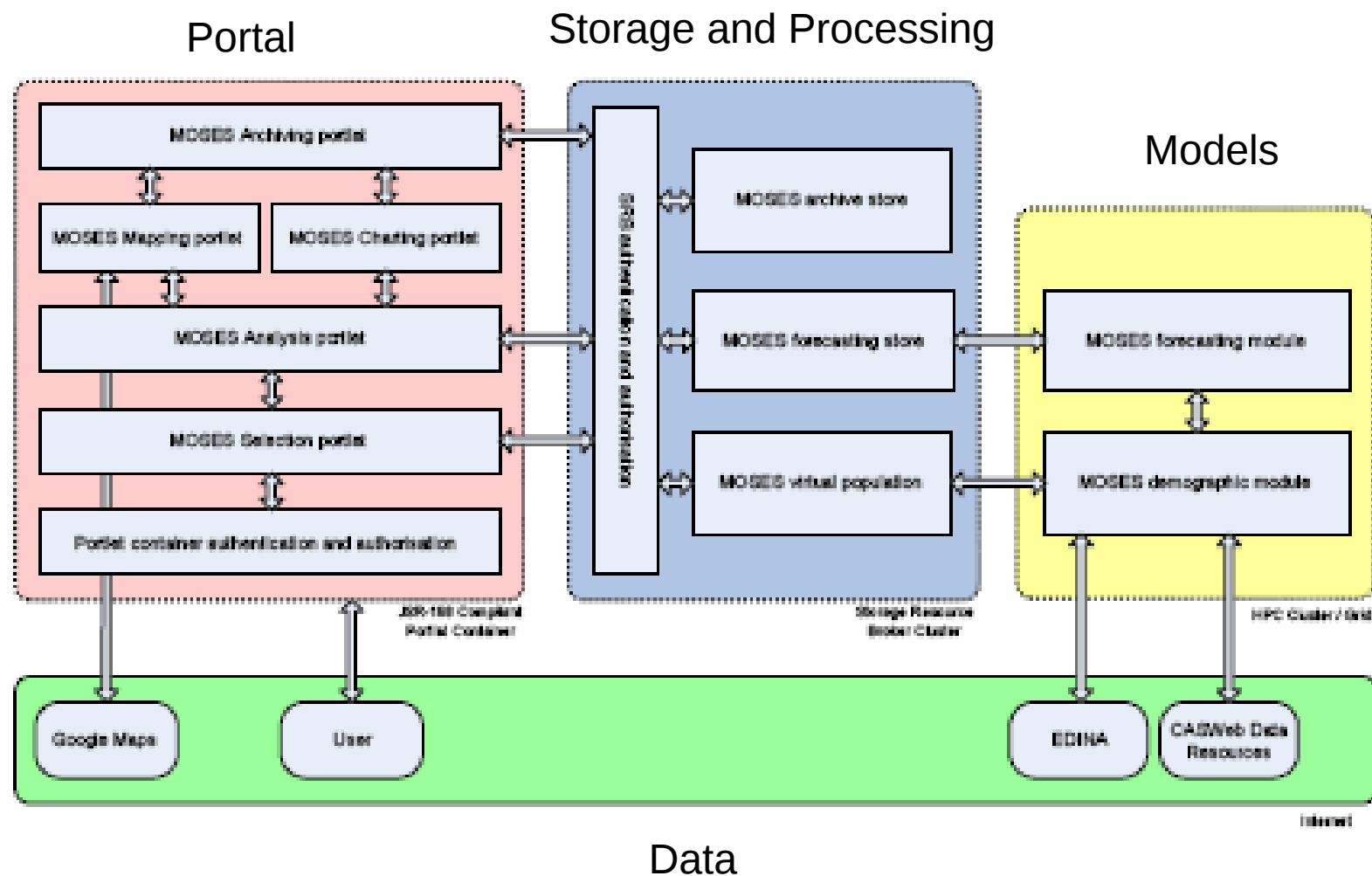
Web portal allows user to select areas of interest and other drop-down functionality



Architecture

- Social simulation is computationally intensive, so that realistic problems are beyond the reach of desktop computers
- Social simulation involves the integration of multiple data sources – access to data can be virtualised within an e-infrastructure framework
- Social simulation may be accessed by multiple remote users, for example by strategists within different regional healthcare providers, or by planners in the social services and housing teams within the same local authority.

Architecture



Next Steps

In the next phase of development, Moses will become more closely aligned with the GeoVue node of NCeSS.

Working together through the **Genesis** programme, the two nodes will begin to prepare a platform providing the methods, data and infrastructure for generative social science.

Through further development of the analytical methods and visual tools supporting Moses, we aim to promote continued awareness of e-social science in both academic research and the practice of public policy simulation.

Generative Social Science

Simulation of artificial societies *in silico* from the bottom up

Evolutionary systems theory combining elements of game theory, experimental economics, complexity and institutional analysis

Implementation via grid-enabled computation and data resources