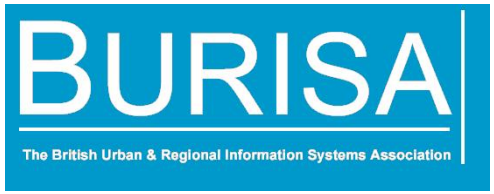
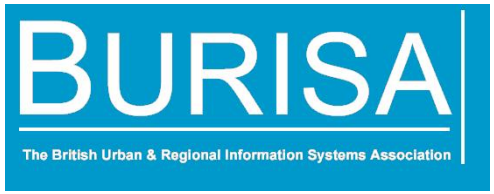
Geographical referencing: A practical workshop

A joint Geo-Refer and SASPAC training event
City University London, 9 December 2008



Welcome and introductions

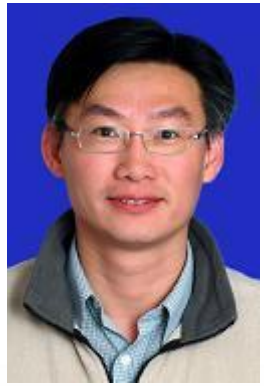
David Martin



Workshop contributors

- David Martin
- Samantha Cockings
- Samuel Leung
- Alan Lewis






- URLs

  http://www.geog.soton.ac.uk/geo-refer/index.html

  http://www.saspac.org

The Geo-Refer Projects

- ESRC Researcher Development Initiative
- 2 Geo-Refer projects
- Focus on training resources in geographical referencing issues
- Particularly aimed at users who are not geographers/ geographical information specialists
- Current emphasis on census, local government and health communities

SASPAC

- Small area Statistics package
- Owned and managed by the Public Sector and is '**not for profit**'
- Funded purely through user's annual subscription fees
- Peak membership over 380 organisations (inc. local, central and regional gov't, health sector, academia, Census Offices, private sector)
- Strong association with the Census (1981-2001) - developed by Census data users for Census data users

Purpose of the workshop

- Conceptual understanding of issues involved in geographical referencing of social, economic and environmental datasets
- Practical skills and methods, including some key datasets – hands-on session with own data or examples provided
- Share questions and experiences with contributors and participants

Schedule of the day

09:30 Arrival and coffee

10:00 Welcome and introductions – David Martin

10:20 Geo-referencing concepts and methods – David Martin and Samantha Cockings

11:30 Break

11:50 Geographical referencing using SASPAC – Alan Lewis

13:00 Lunch

13:45 Briefing for hands-on workshop session

14:00 Hands-on session (participants' may use own datasets)

16:00 Summary and overview of lessons learned

16:15 Close

Rules of engagement

- This is an informal workshop
- There is no such thing as a silly question!
- Please interrupt

Georeferencing examples

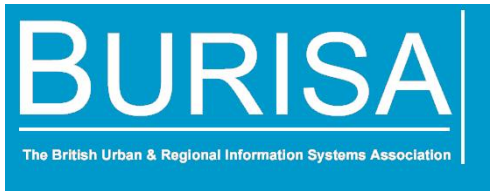
- Link survey results to census data
- Relate local services to indices of deprivation or area classifications
- Aggregate and map local data with national Neighbourhood Statistics
- Identify which service delivery locations fall within different policy areas

Why georeference?

- Massive growth in geographical data
- Adds analytical value
- Fundamentally, geographical referencing leads to either:
  - Data linkage, potentially for aspatial analysis
  - Mapping, and other forms of spatial analysis

Geo-referencing concepts and methods

David Martin, Samantha Cockings



Geographical referencing of social science phenomena: 1

- Characteristics of people or events, recorded at mail addresses
- Geographical location of mail address

Geographical referencing of social science phenomena: 2

- Administration or policy related to defined areas, aggregations of individuals
- Geographical location of area boundaries

Geographical referencing of social science phenomena: 3

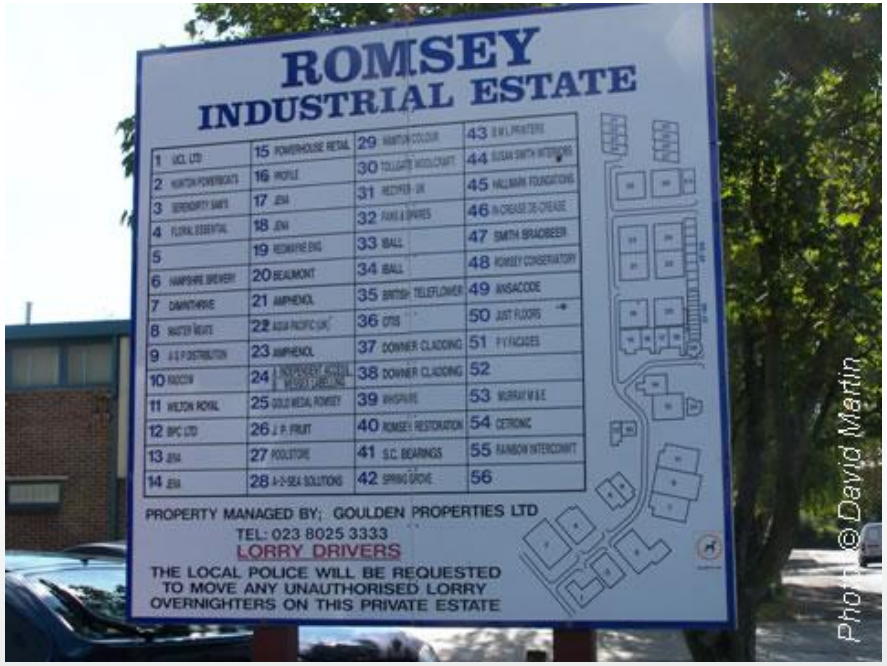
- Phenomena that are linear, describing routes or flows
- e.g. bus route, telephone conversation, social network link

Geographical referencing of social science phenomena: 4

- Phenomena or events that have locations but do not fit standard descriptions
- e.g. road accident, environmental quality

Geographical referencing of social science phenomena: 5

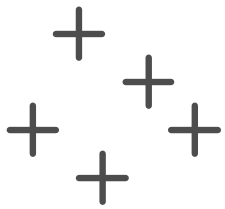
- Complex phenomena not captured by a simple location
- e.g. business locations, catchment populations



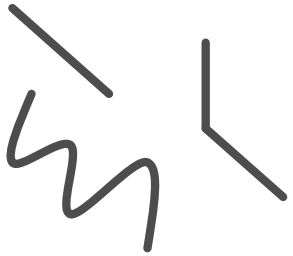
Geographical object types

POINTS

E.g.
Health data
Meteorological data
Pollution sources
Hospitals
Landfill sites

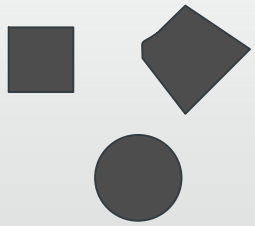


LINES



E.g.
Roads
Powerlines
Railways
Rivers

AREAS



E.g.
Enumeration Districts
Wards
Health Authorities
Grid squares
Water supply zones

+ NETWORKS?
+ SURFACES?

Geographical object types

- **Points: a single pair of coordinates**
- Lines: an ordered sequence of coordinates
- **Areas: closed, ordered sequences of coordinates**
- Networks: complex line-based structures
- Surfaces: mathematically modelled representations of continuously varying phenomena

Direct and indirect georeferencing

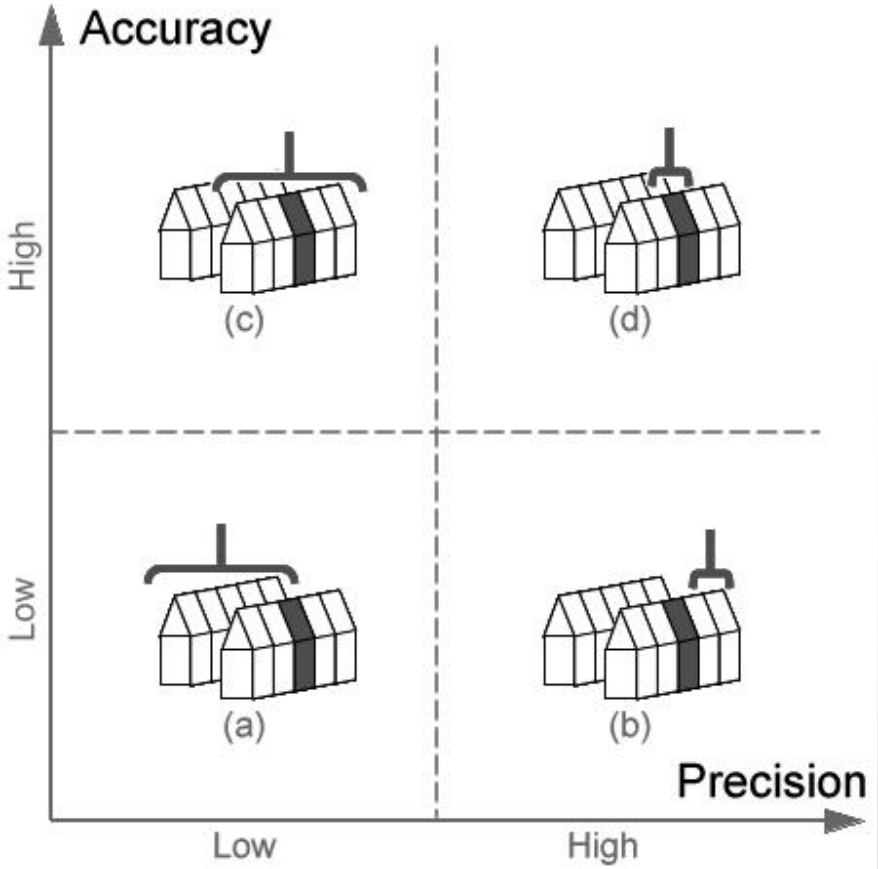
- Direct georeferencing: explicit coordinate system
  - Lat/Long, Ordnance Survey grid reference
- Indirect georeferencing: any type of area code or name relating to a known location
  - Postcode, Zip code, County, Census output area/enumeration district/tract; local government district; health authority, etc…

Where do coordinates come from?

- Surveys for mapping
- Remote sensing
- Digitising paper source documents
- Direct capture using Global Positioning System (GPS) receiver
- **Generally contributing to reference datasets created for re-use**

Accuracy and precision

- Accurate location
  - free from locational bias
  - e.g. respondent lives in Southampton
- Precise location
  - provides detail (but not necessarily accurate)
  - e.g. respondent lives at 25b Highfield Lane
- Need to assess and adopt different strategies according to purpose

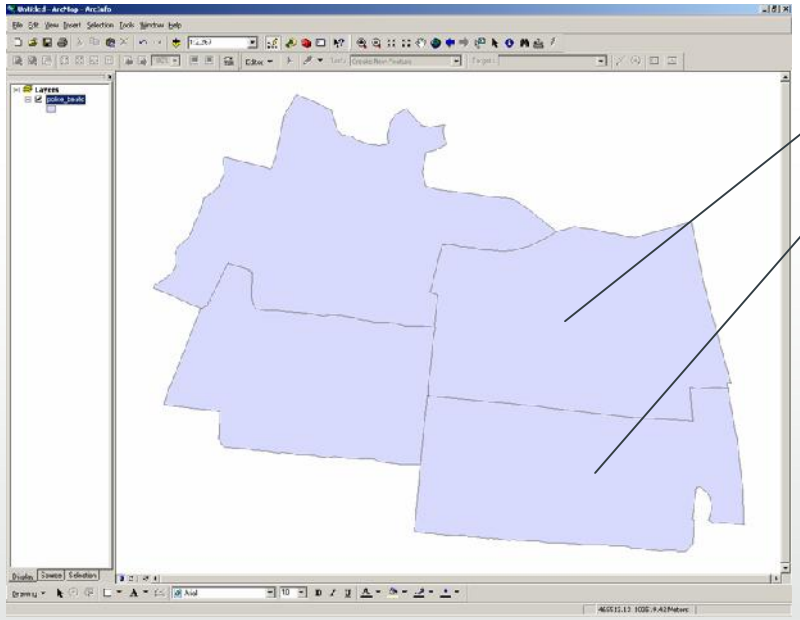


Scale and projection

- Map scale
  - Explorer map: 1:25,000
  - Road atlas: typically 1:400,000
- Leads to inclusion/generalization of different features, boundary details,etc.
- Projection systems:
  - Very important when using lat/long datasets,
  - GB National Grid sidesteps the issues…

Locations and attributes

- Locations: points, area boundaries



Attributes of BEATATTS

OID	BEATID	NAME	NOFFS	AUTH
0	1	Westend	30	Anywhere
1	2	Southend	20	Anywhere
2	3	Eastend	5	Nowhere
3	4	Northend	10	Nowhere

Record: 1 Show: All Selected Records (0 o

- Attributes: values, characteristics

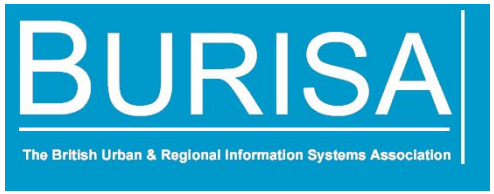
Geoprivacy issues

- Detailed location is potentially disclosive
- Postcode is generally considered to be sensitive information
- Explicit subject consent in data collection?
- Subject identification may be possible through linkage and mapping
- All usual rules apply!

What about GIS?

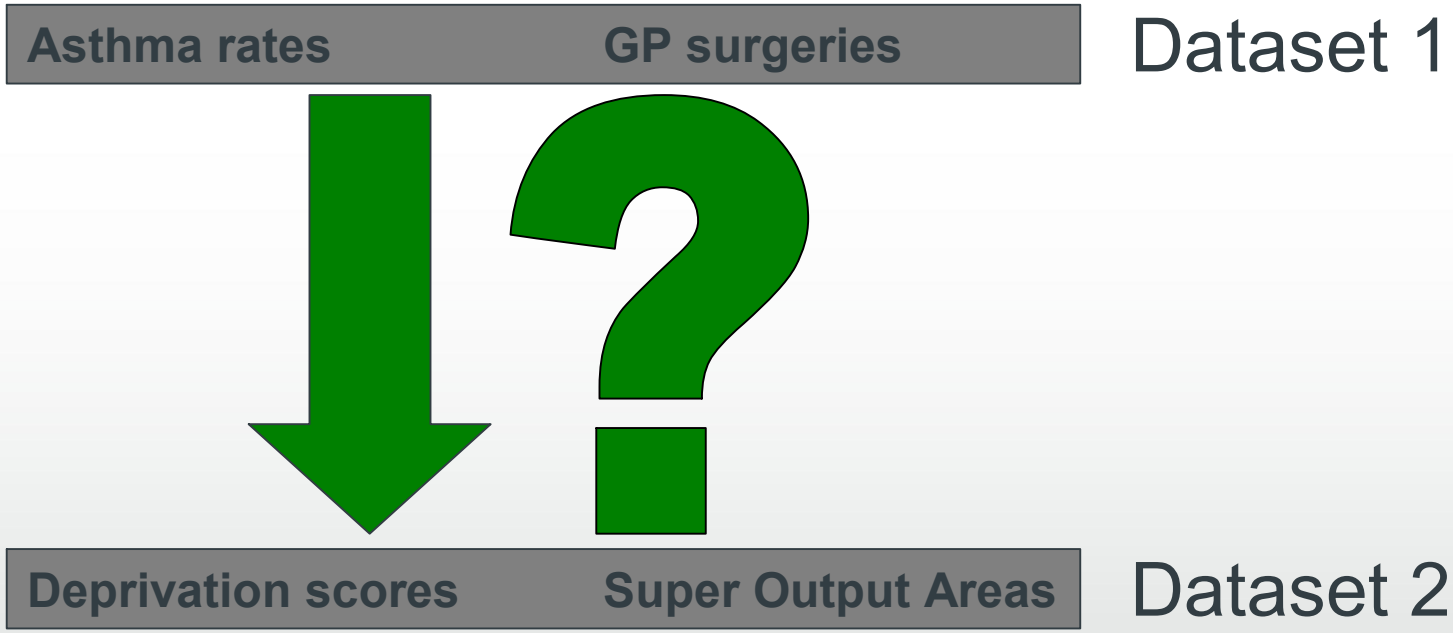
- Massive growth in use of GIS since early 1980s: complex information systems
- GIS growth promoting data standards, growth of geoinformation industry etc.
- GIS provide useful georeferencing tools and lots of other functions not needed here…
- MapShore mapping system within SASPAC
- This is not a GIS course!

Georeferencing methods: linking data



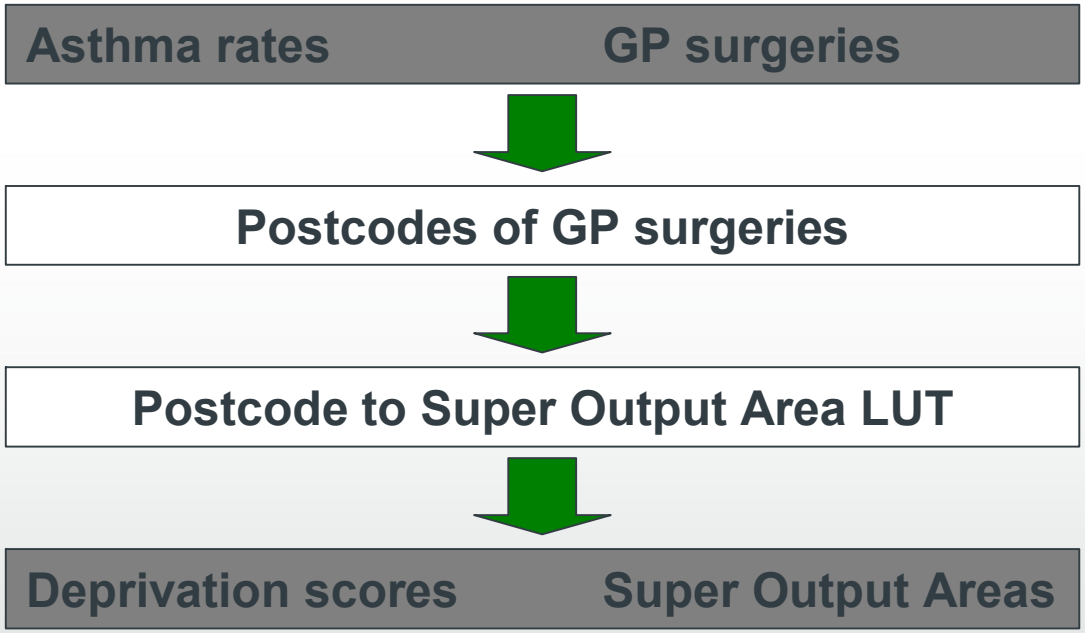
Tabular data linkage example

Research question: What is the relationship between asthma and deprivation?



Tabular data linkage example

Tabular linkage requires lookup(s) between georeferences on two or more datasets



Dataset 1

Georeferences

Geographical lookup table (LUT)

Dataset 2

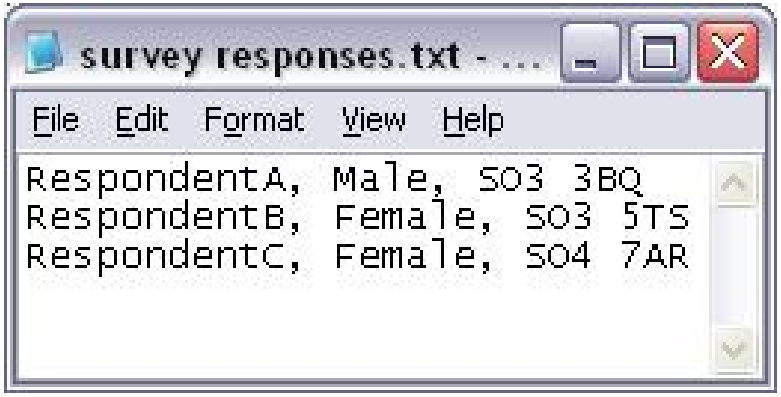
Assumes surgery is appropriate georeference for patient data!

Tabular linkage tools

- Access, SPSS etc. – general purpose database or statistical packages for matching lists with common data items
- SASPAC provides facilities for data aggregation from gazetteer files, e.g. from census OAs to locally-defined neighbourhoods

Data linkage example: input tables

Coded survey responses



Respondent ID

Gender

Postcode

Postcode directory

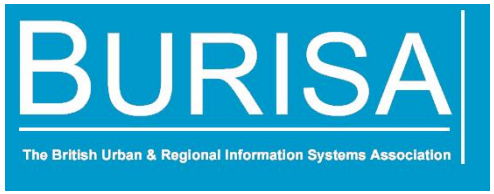
Postcode

Ward code

Watch the time!

- Extreme care needs to be taken when matching between datasets relating to different dates
- Postcodes, boundaries, area names all subject to change
- All changes on different timescales
- Assume another time = another geography!

Georeferencing methods: mapping data



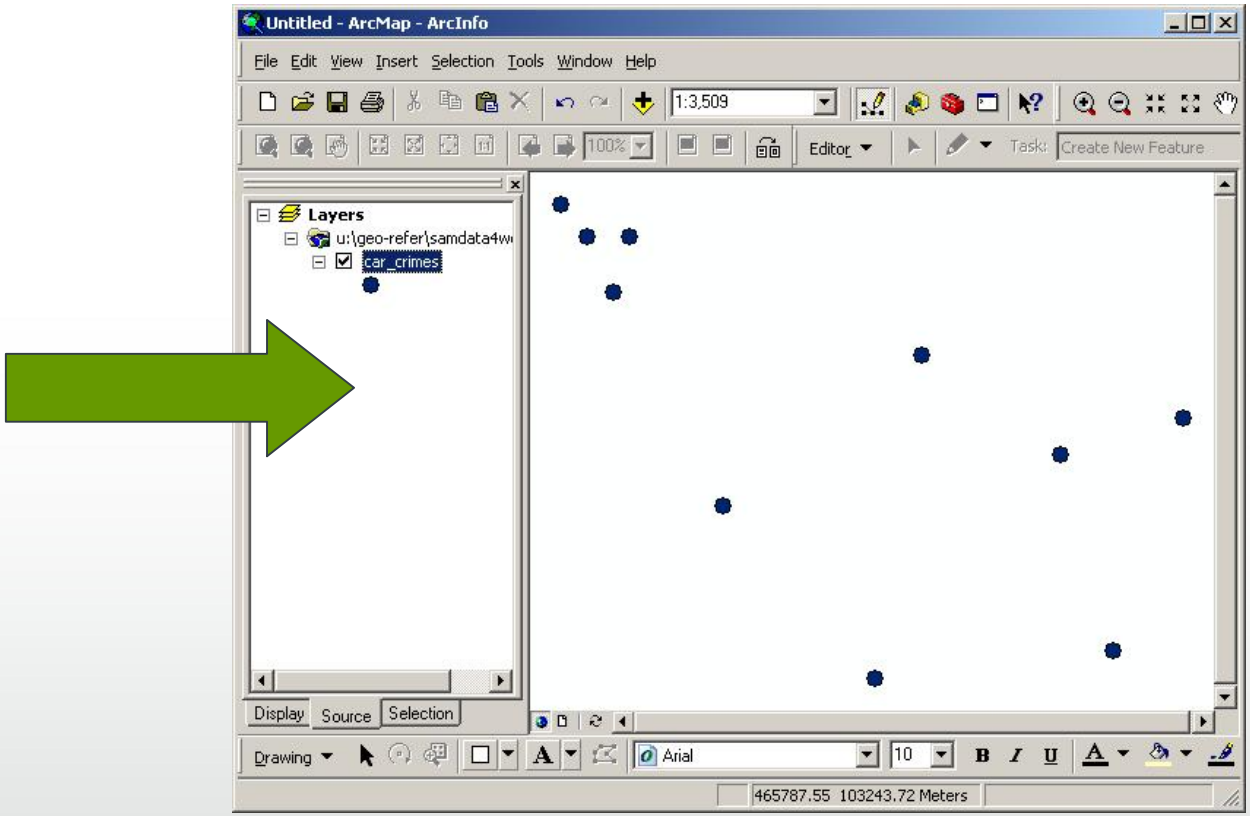
Mapping

- Map locations of points and/or areas
  - e.g. grid-refs of survey responses
  - e.g. boundaries of Lower Layer Super Output Areas (LSOAs)
- Map attributes of areas
  - e.g. Link survey responses on attitudes to recycling to wards, then map response rates by ward

Mapping grid-referenced points

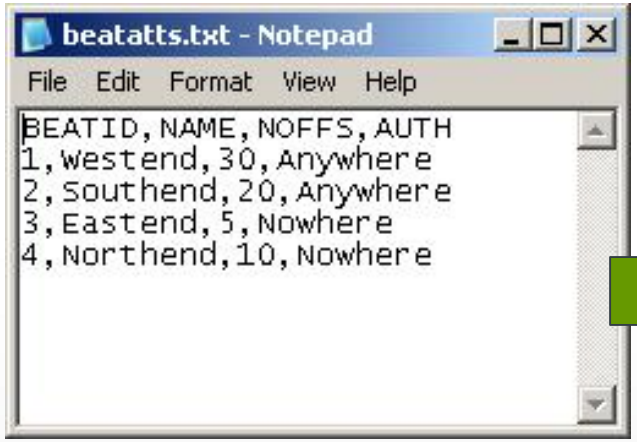
Grid-refs of car crimes

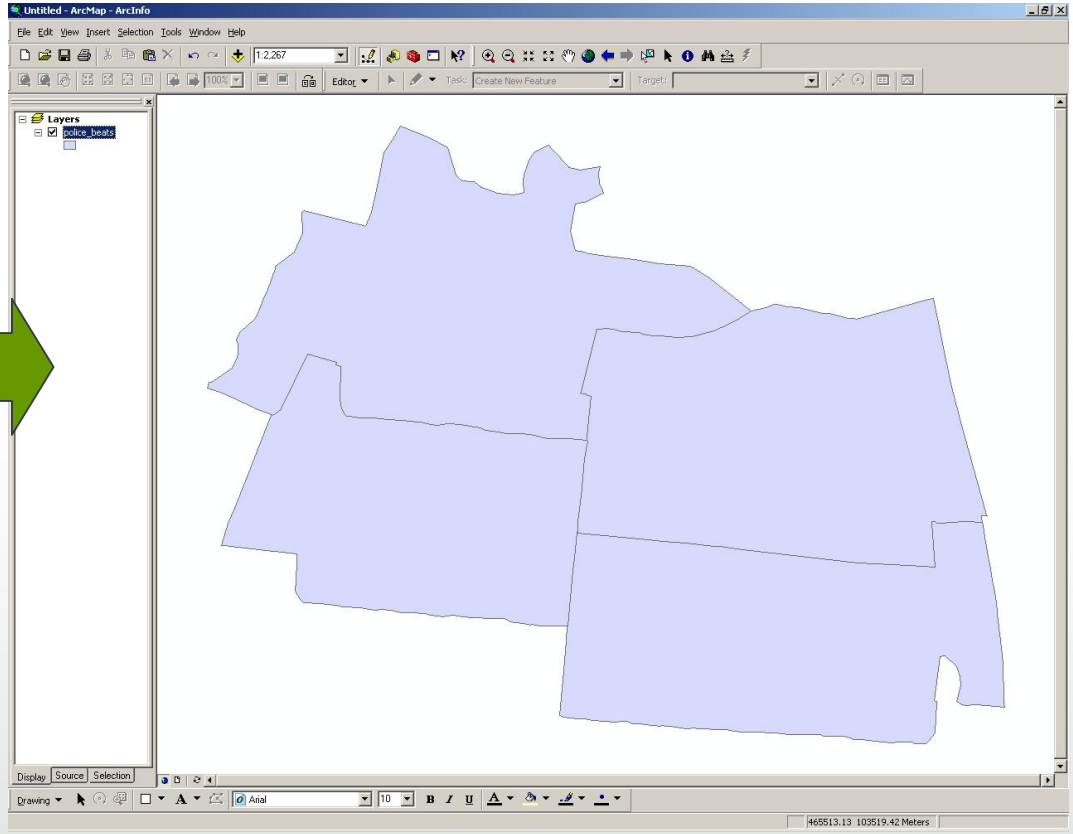
465474,103429
465490,103409
465517,103409
465507,103374
465703,103334
465791,103271
465869,103294
465577,103238
465825,103146
465673,103129



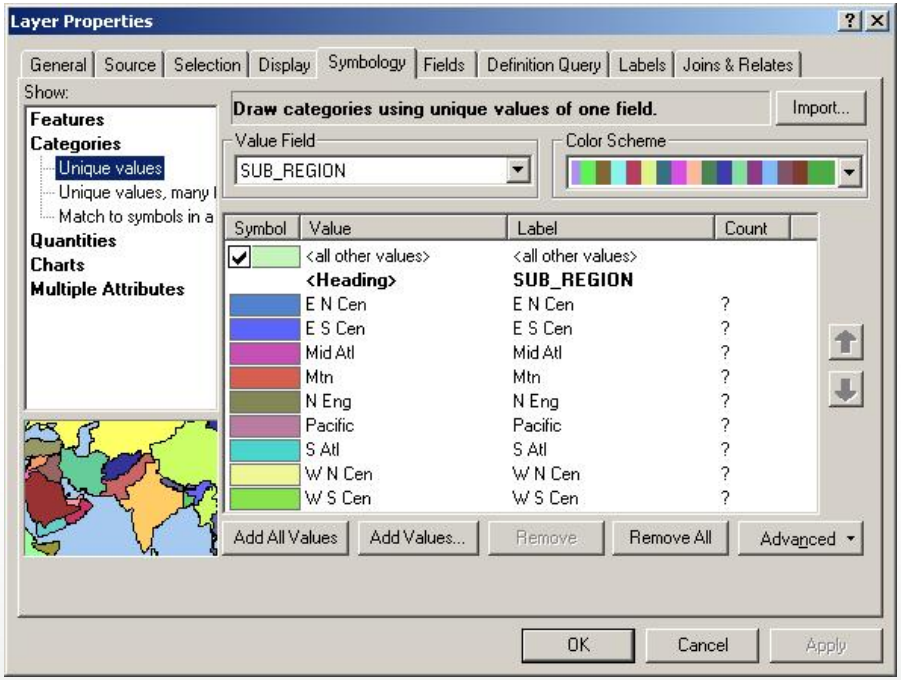
Linking attribute data to boundary data

Area-based attribute data

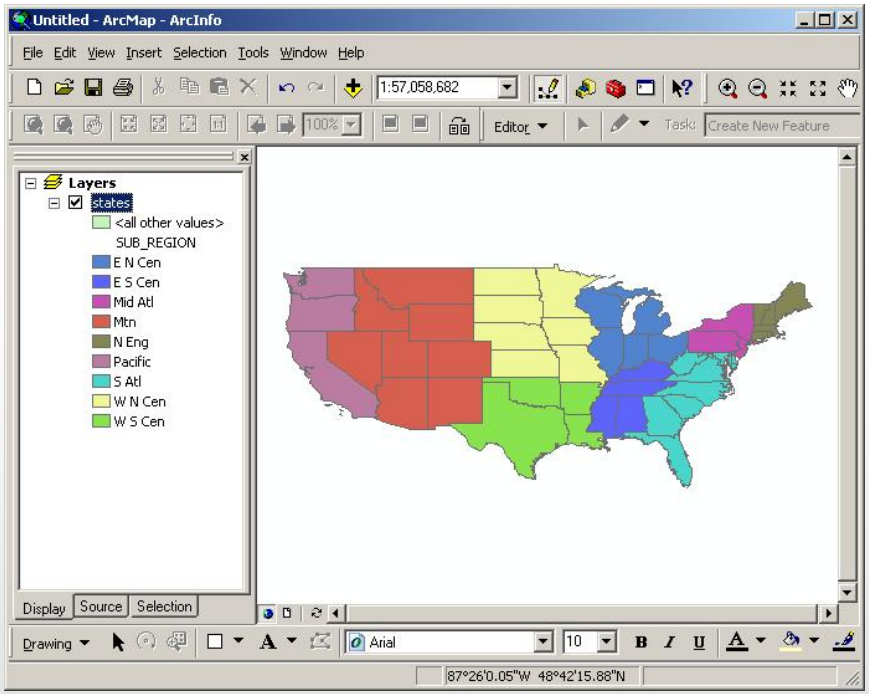




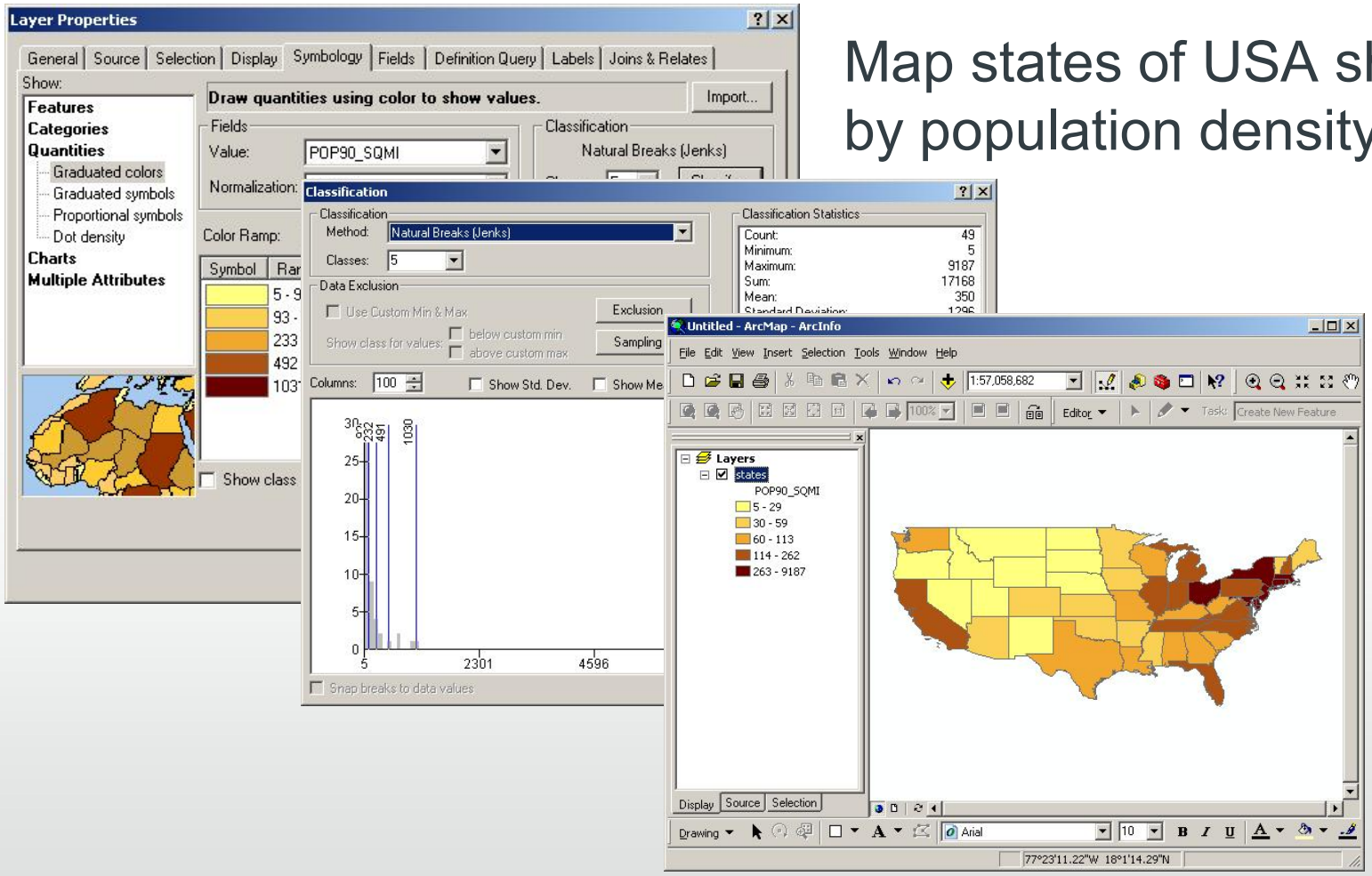
Mapping area data: Categorical



Map states of USA shaded by sub-region



Mapping area data: Continuous



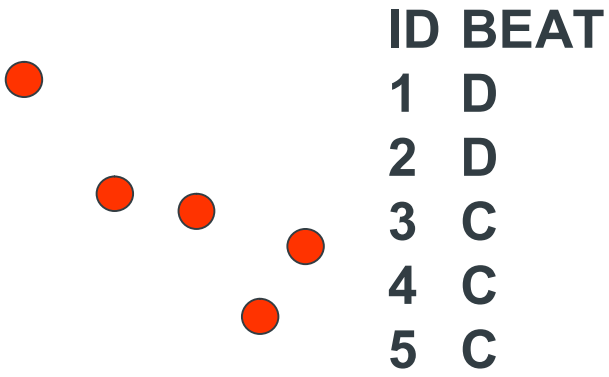
Map states of USA shaded by population density

Spatial data linkage

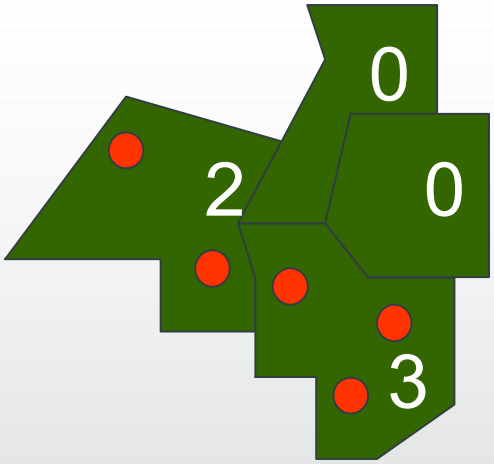
- Where no tabular linkage exists between data
- Can use GIS operations to match points to areas or areas to areas
  - e.g. grid referenced accidents matched to own neighbourhood boundaries
  - e.g. clinic catchment areas to new ward boundaries

Allocation and aggregation

- Allocate
  - e.g. allocate attributes of police beats to car crime locations
- Aggregate
  - e.g. aggregate number of car crimes in each police beat
- Same spatial linkage operation, but different outputs



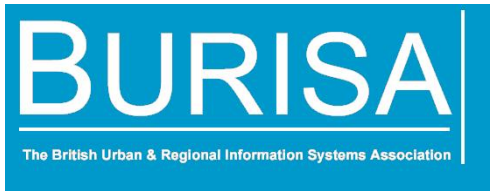
ID	BEAT
1	D
2	D
3	C
4	C
5	C



Modifiable areal unit problem

- Pattern observed in shaded area census maps are artifact of boundaries used
  - Different boundaries would produce different pattern, even with same population
  - Relationships at one level of aggregation need not hold at any other level (ecological fallacy)

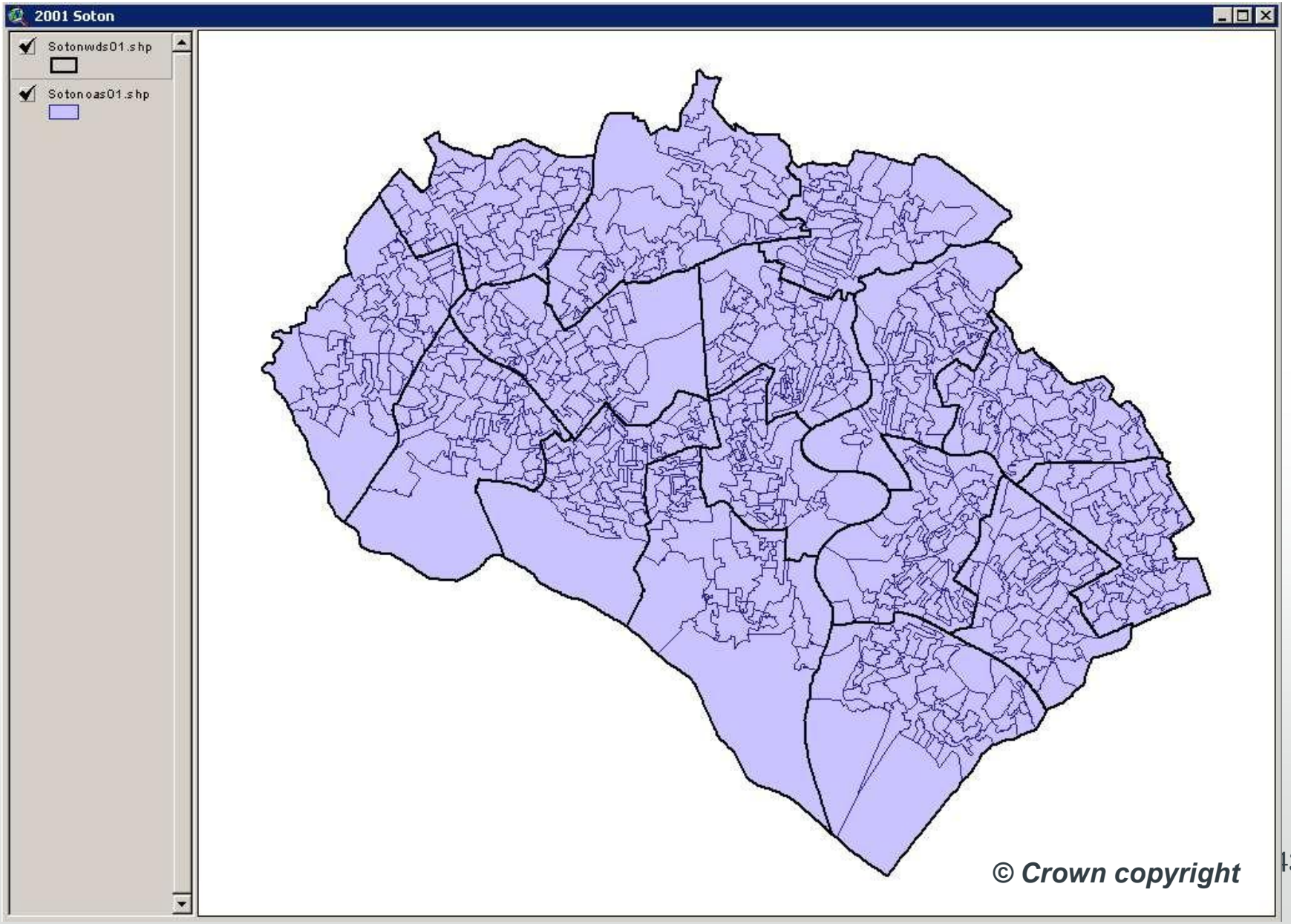
Implications for some commonly-used datasets



Census geography: codes and boundaries

- Area hierarchy and codes
  - e.g. 25JPUF0005
  - Output Area (no names) 0005
  - Ward UF
  - LADs (county/district 25JP or unitary authority 00JP)
- Digital boundaries
  - Available all levels from OA upwards

Southampton: 2001 output areas and wards



Census geography: relationships

- Relationship to other geographies
  - OAs mostly built from unit postcodes
  - OAs nest within LSOAs
- Change over time
  - No direct relationship between 2001 and previous censuses
  - Wards (and local authorities) change between censuses

Census geography: future

- ONS and University of Southampton development work underway
- Basic principle of stability, from user consultation
- Some essential splits and mergers to accommodate inter-censal population change
- Very limited redesign due to 2001 census deficiencies
- Aim is for revisions to be nested and geographically constrained

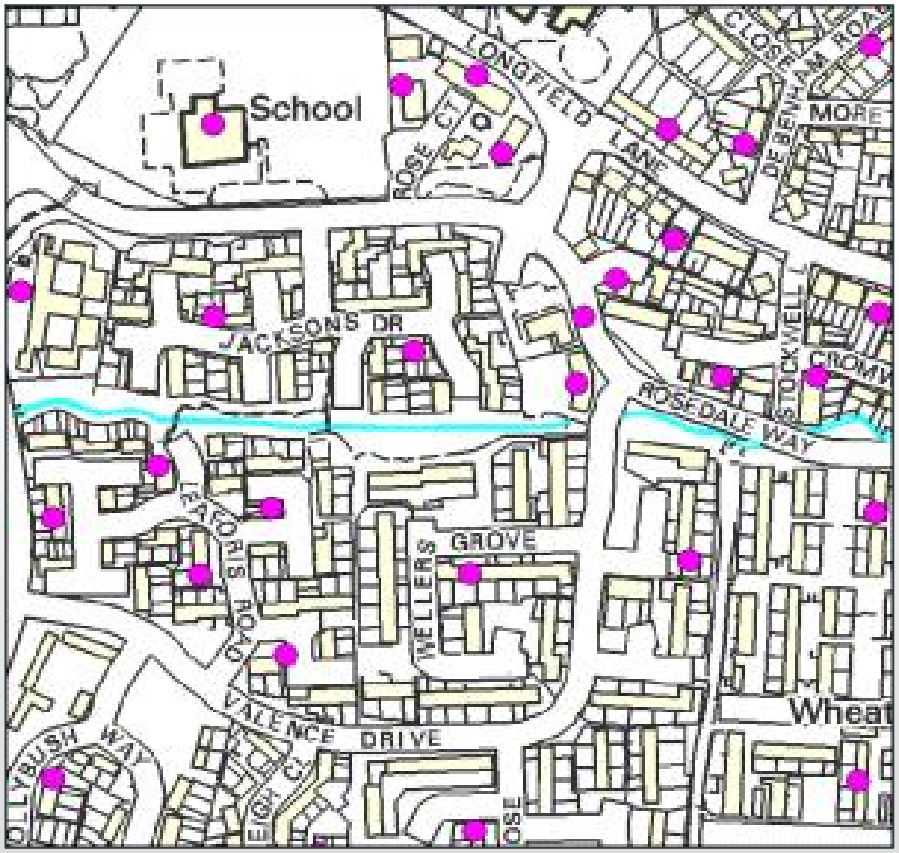
Postcodes: coding structure

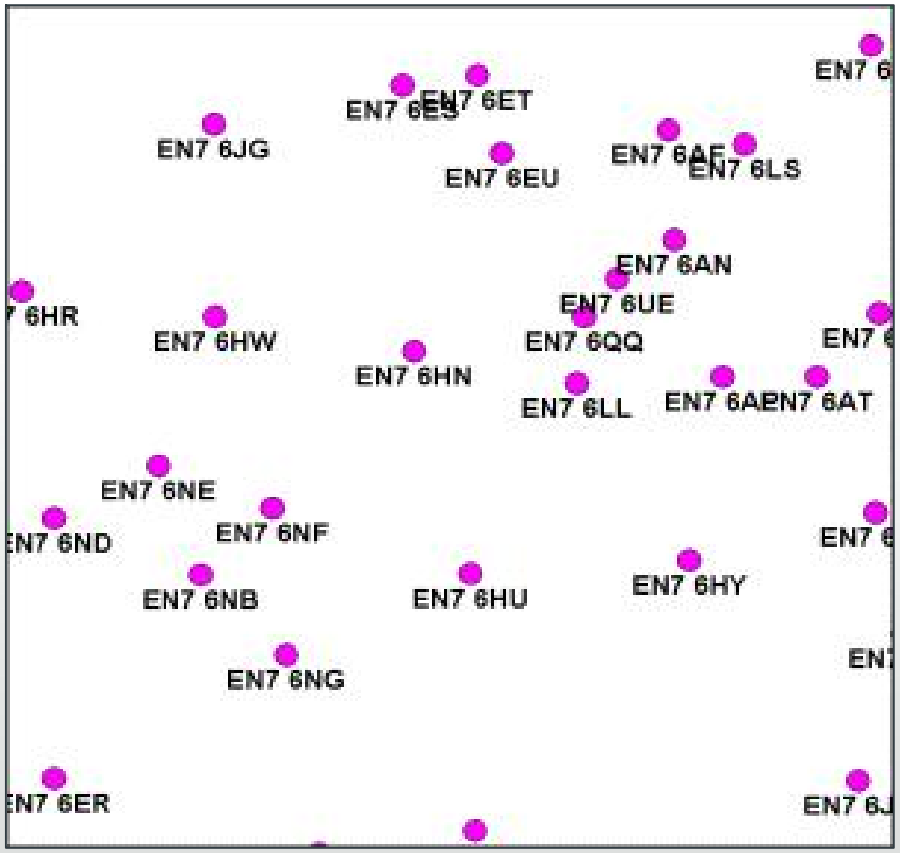
- Area hierarchy and codes
  - Unit postcode (NOT areas) BS8 1SS
  - Sector BS8 1
  - District BS8
  - Area BS
- Large/small users/non-geographic codes

Postcodes: spatial referencing

e.g. Ordnance Survey Code-Point

http://www.ordnancesurvey.co.uk/oswebsite/products/codepoint/





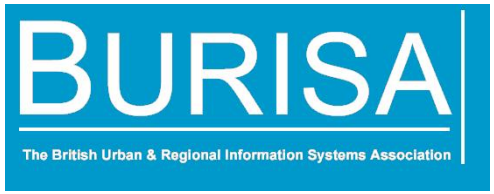
Postcodes: relationships

- Relationship to other geographies
  - Controlled directly by Royal Mail
  - No direct relationship to any other geographical units
- Change over time
  - Continuous small-scale change
  - Periodic large-scale reallocation
  - Eventual reuse of discontinued codes

National Statistics Postcode Directory (NSPD): what is it?

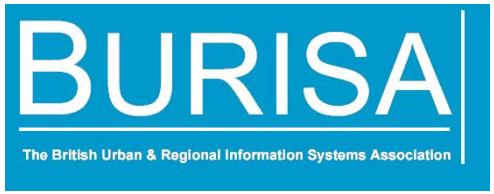
- A long list of UK postcodes
- Produced by ONS, Royal Mail and Ordnance Survey
- Multiple additional geographical codes
- Postcode metadata
- Updated quarterly

Morning break!!!



Geographical referencing using SASPAC

Alan Lewis



Overview:

- Introduction to SASPAC
- Software features
- Accessing data (worked example 1)
- Data aggregation/rezoning (worked example 2)
- Mapping data in MapShore (GIS) (worked example 3)

Introduction to SASPAC: 1

- Software to provide link between the complex Census data provided by the Census Offices and the user
- Store, manipulate, export and publish Small Area Statistics (LAs, wards, Output Areas, SOAs, Postcodes)
- Desktop, network, thin-client application
- Owned by IDeA and managed by GLA

- '**Not for profit**' initiative

Introduction to SASPAC: 2

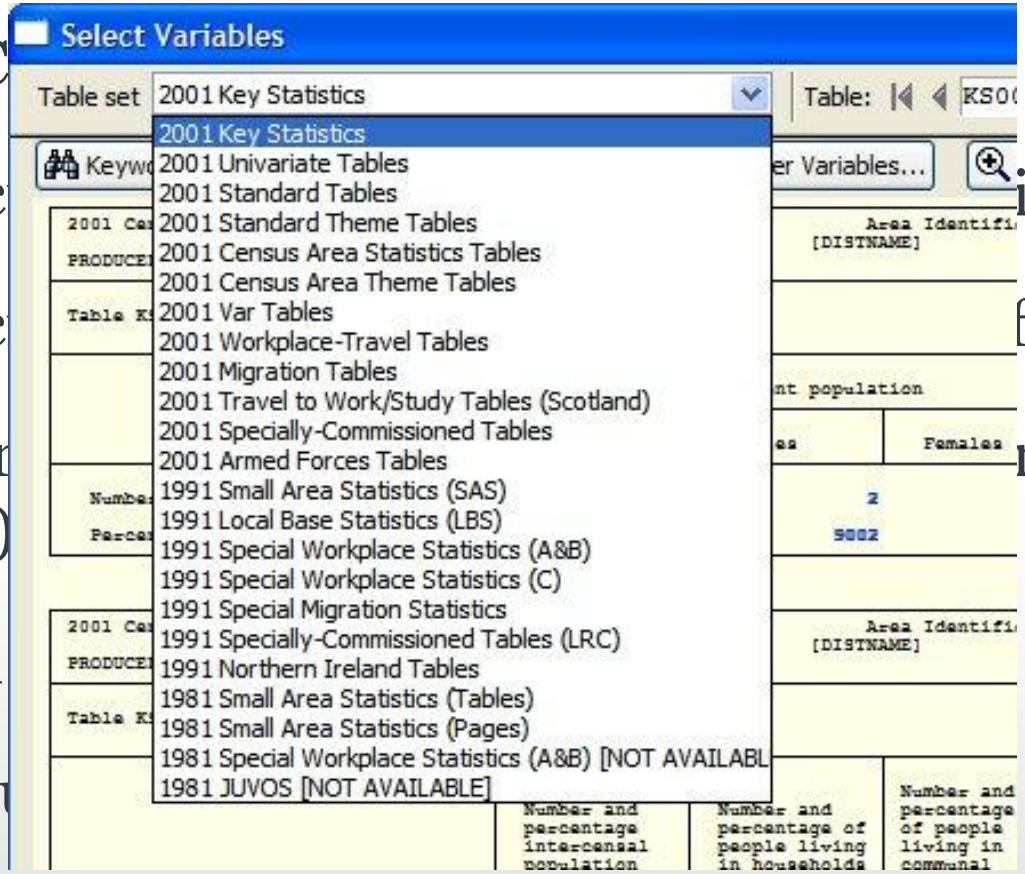
- Peak at almost 400 organisations following the release of 2001 data
- UK – wide user group, includes: local, central & regional gov., Health sector, Census Offices, Academia, Private sector
- Site licence is £1350+VAT and annual Support fee is £600+VAT
- Developed for the users by the users - SASPAC Advisory Panel (SAP)

Introduction to SASPAC: 3

- Software releases – current v9.00 Nov'08
- Software maintenance/upgrades/enhancements
- Historical and new datasets
- Helpdesk support/advice (web, email, phone)
- Training

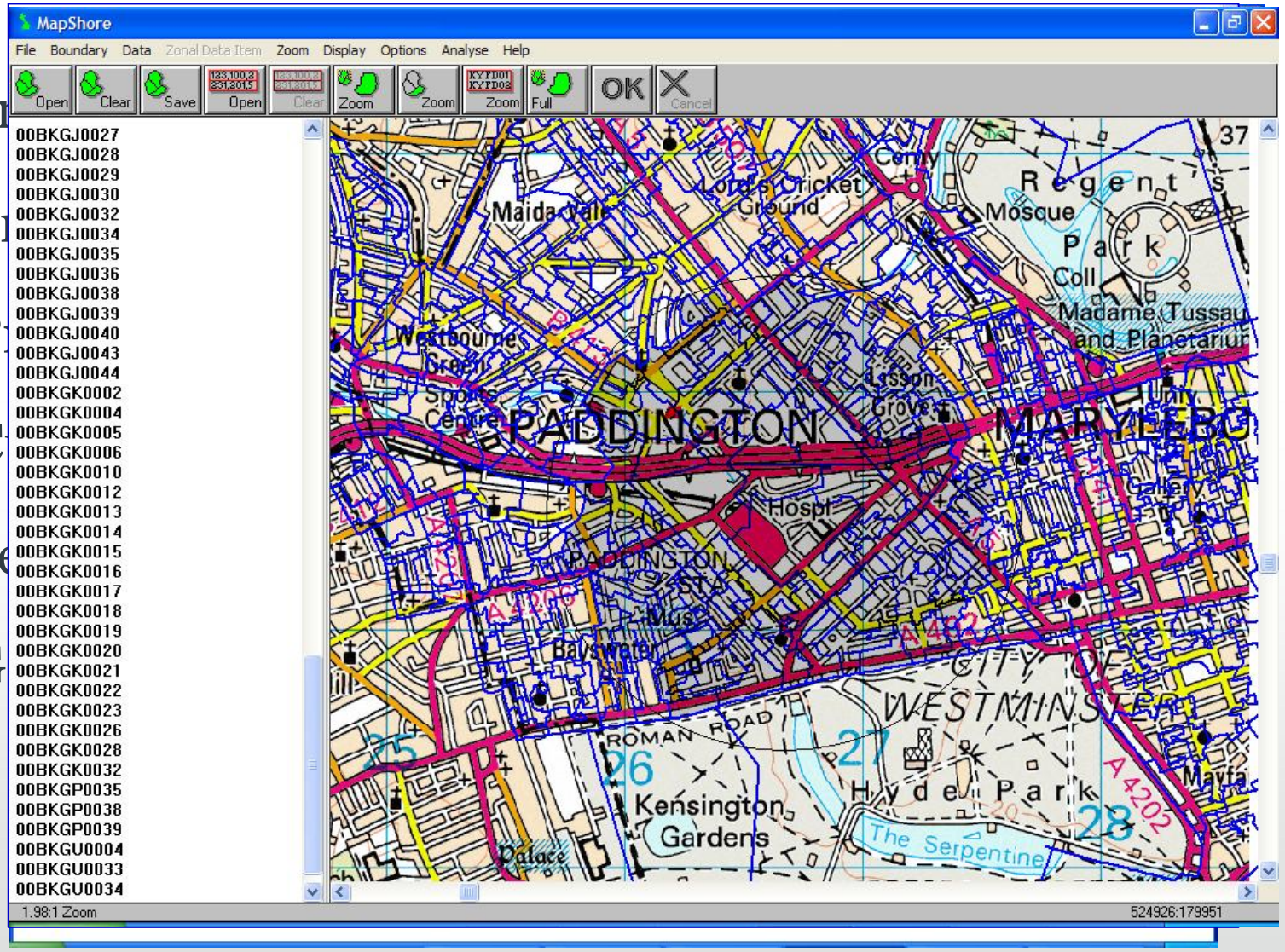
Software features: 1

- All 2001 C
- 2001 Spec ics
- 2001 Spec far!)
- New geogr rliamentary Cons…etc)
- 1981/1991
- Non-Censu



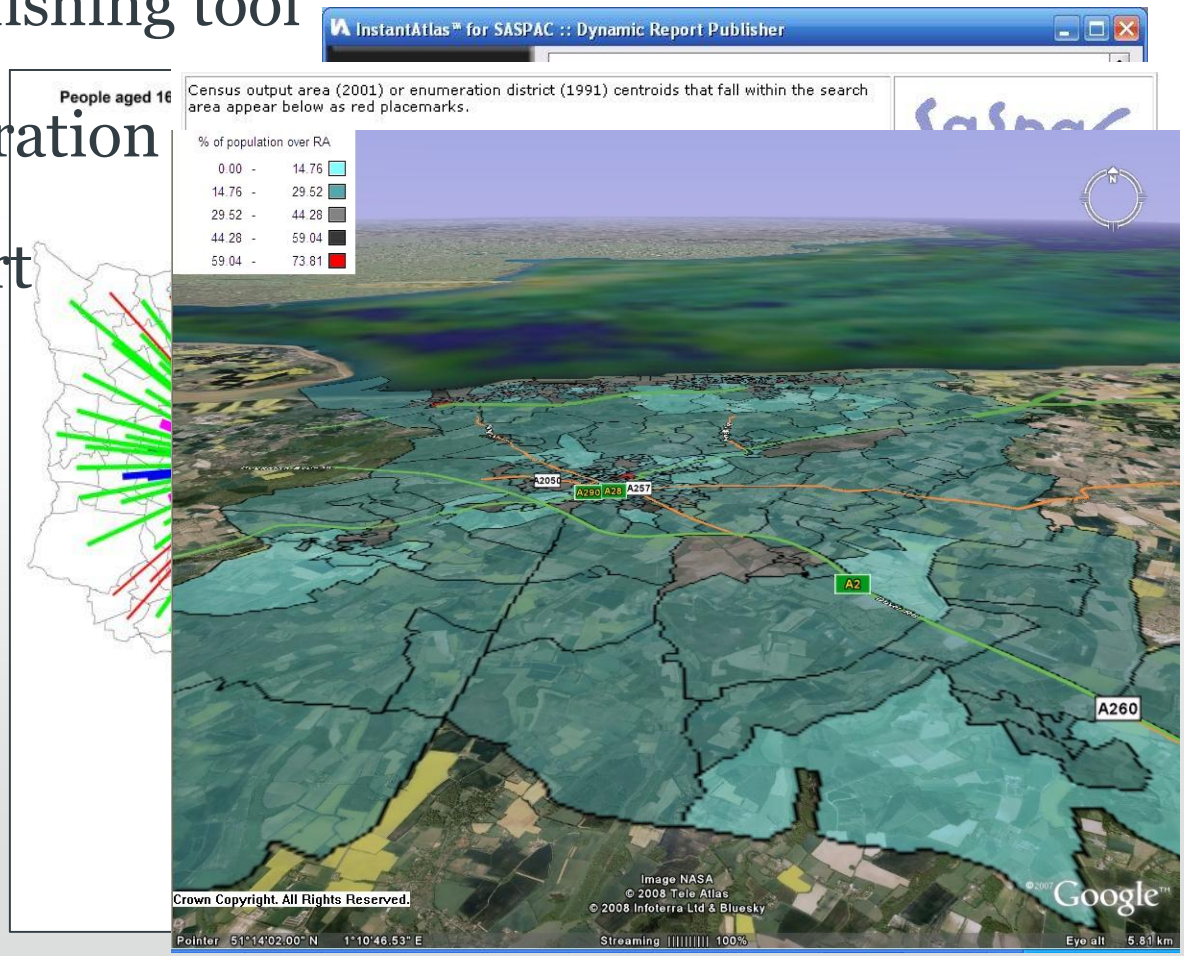
Software features: 2

- br
- Pr
- Py
- C
- Se
- G



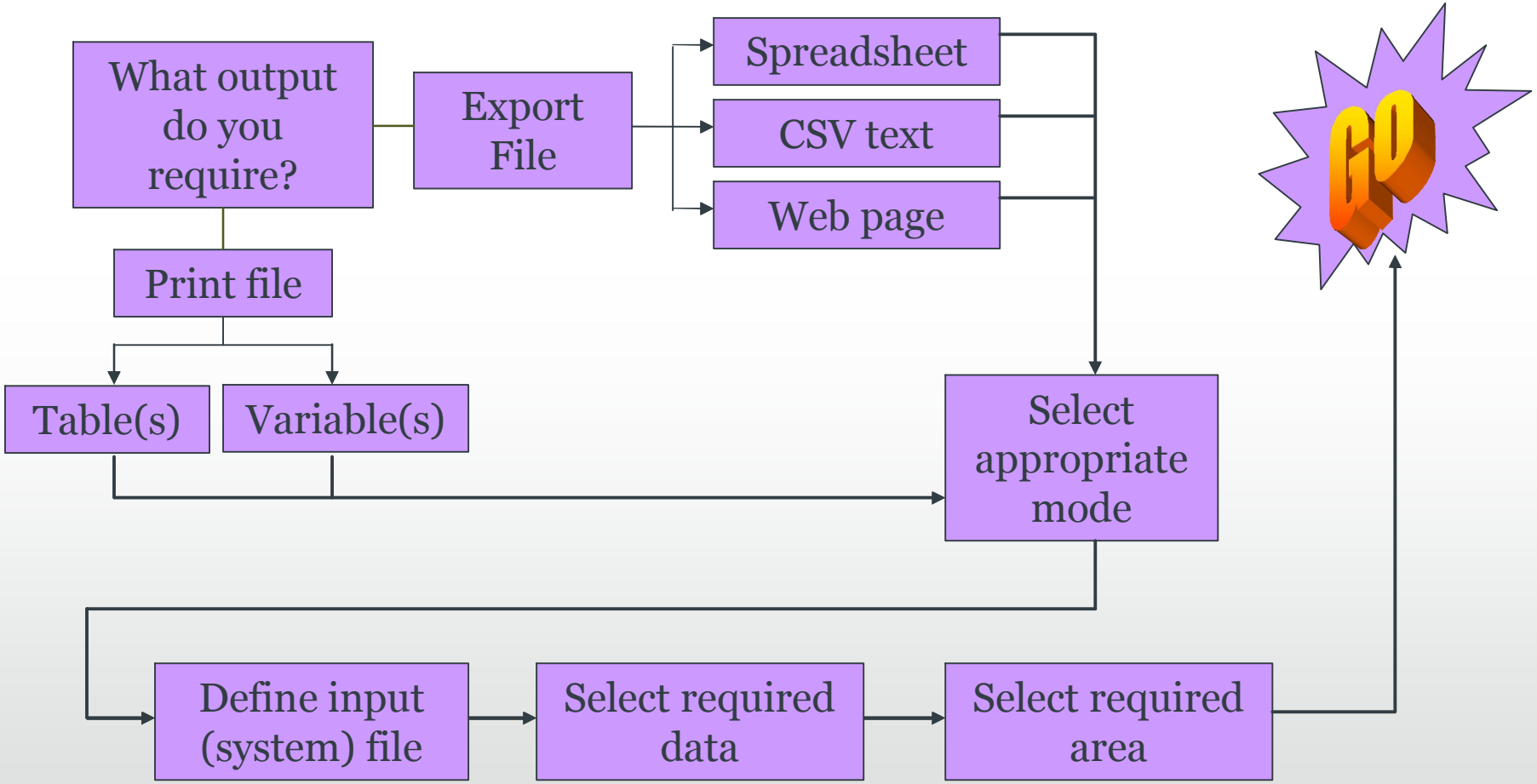
Software features: 3

- Integrated GIS - MapShore
- Integrated Web Publishing tool
- Google™ Maps integration
- Google™ Earth export



Accessing data: 1

- SASPAC task sequence:



Accessing data: 2

Concepts:

- Windo
- Com
- Syst
- Fran

2001 Census Standard Tables — Area Identifier - **064**

PRODUCED USING SASPAC — **ENGLAND**

Table ST001 AGE BY SEX AND RESIDENT TYPE
Table population: All people

	ALL PEOPLE			Males		
	Total	Household residents	Communal Establishment residents	Total	Household residents	Comm Estal resi
ALL PEOPLE	49,138,831	48,248,150	890,681	23,922,144	23,510,259	
0 to 4	2,926,238	2,923,246	2,992	1,498,073	1,496,435	
0	554,460	553,206	1,254	283,071	282,339	
1	574,428	573,918	510	294,528	294,249	
2	587,635	587,173	462	301,224	301,007	
3	596,726	596,351	375	305,362	305,164	
4	612,989	612,598	391	313,888	313,676	
5 to 9	3,122,529	3,118,999	3,530	1,599,800	1,597,648	
5	604,631	604,270	361	309,690	309,502	
6	608,575	608,217	358	312,461	312,268	
7	625,462	625,038	424	320,308	320,084	
8	630,665	629,806	859	323,034	322,520	
9	653,196	651,668	1,528	334,307	333,274	
10 to 14	3,229,047	3,199,552	29,495	1,653,034	1,634,992	
10	661,291	658,955	2,336	338,724	337,180	
11	646,996	643,167	3,829	331,353	329,037	
12	640,717	634,628	6,089	327,785	324,218	

Accessing data: 3

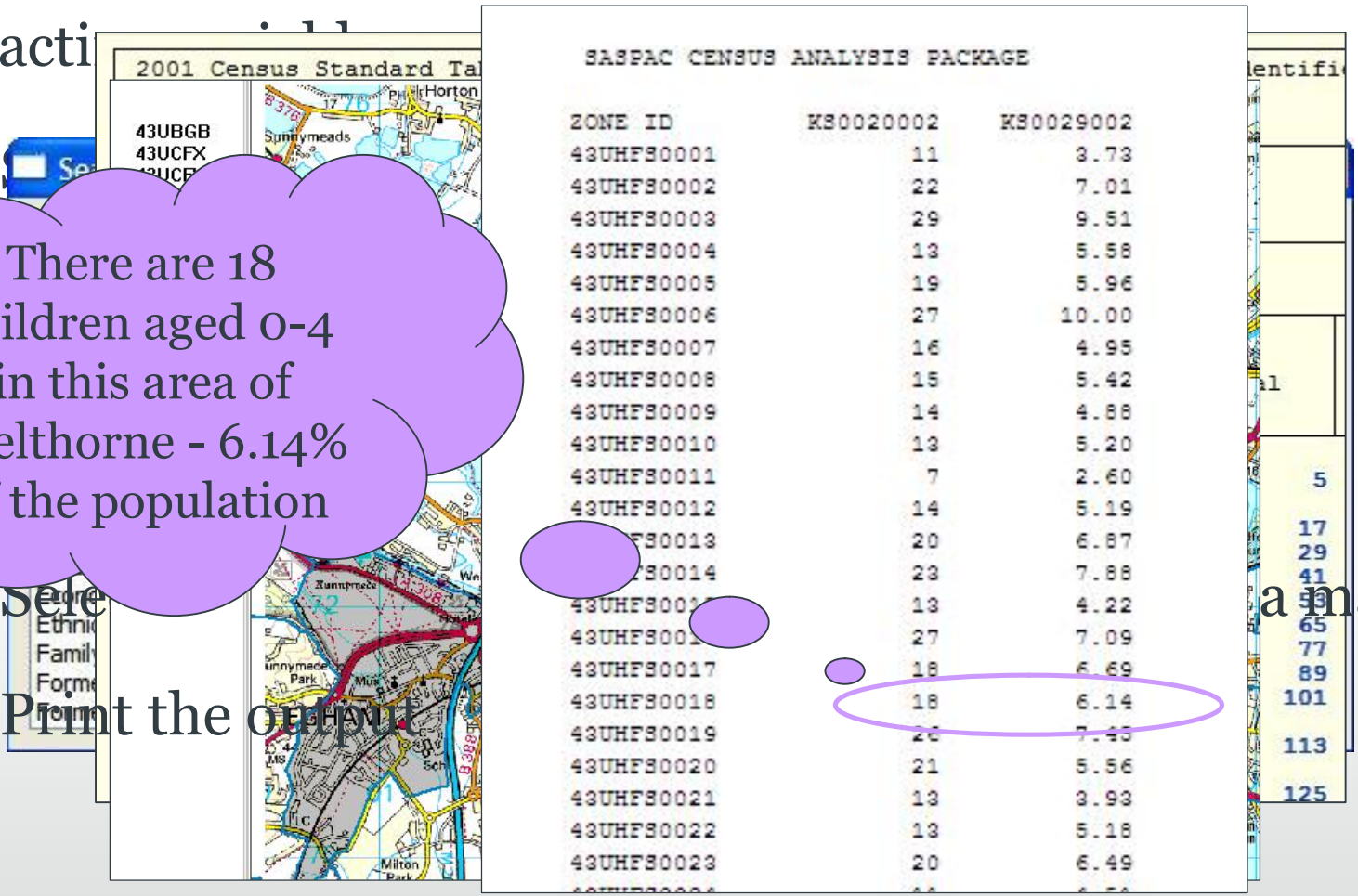
- Extracti

1. S

There are 18 children aged 0-4 in this area of Spelthorne - 6.14% of the population

5. Sele a map

6. Print the output

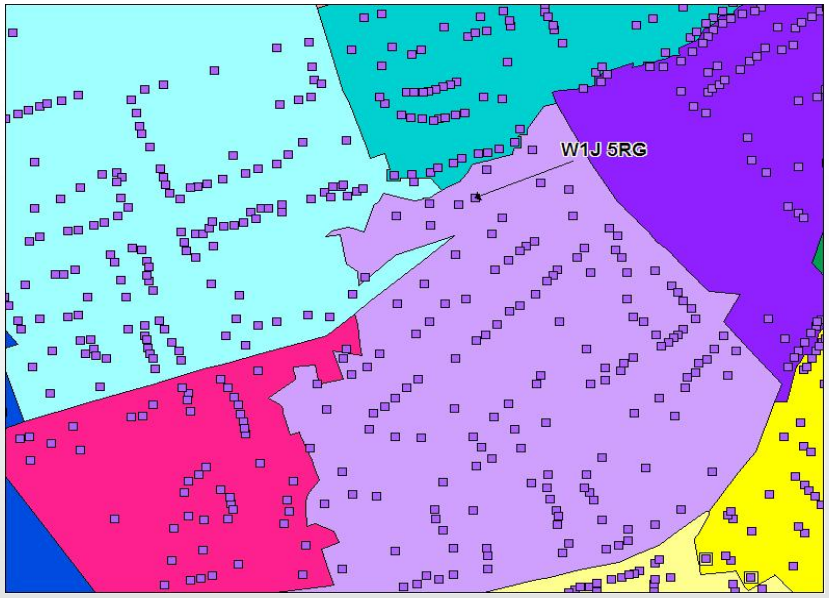


Worked example: 1

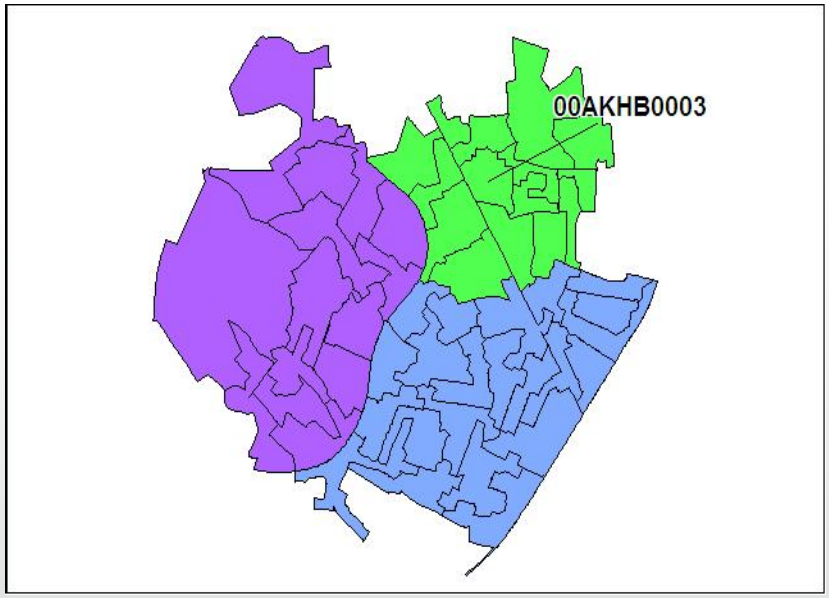
- Print the 2001 Census Key Statistics table relating to country of birth for all the wards in your own Local Authority. Preview the output in SASPAC, Excel and in Internet Explorer as an HTML file.

Data aggregation/rezoning: 1

- Often required to aggregate zones to build new ones
- e.g. police beats, GP catchments, workplace areas
- Use existing points or boundaries as building blocks



Unit Postcodes - Sectors



Output Area aggregation

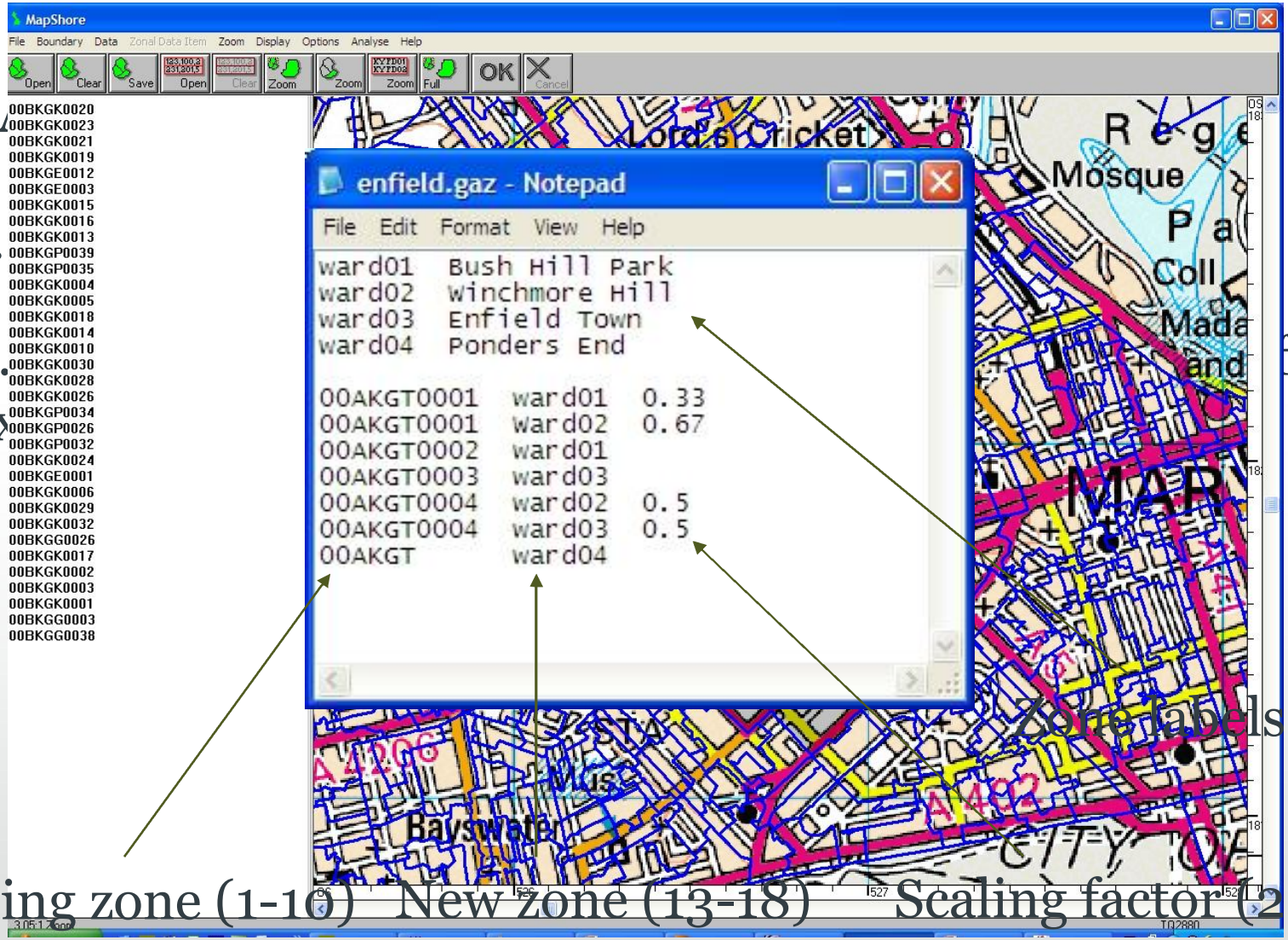
Data aggregation/rezoning: 2

- S

1.

2.

ex



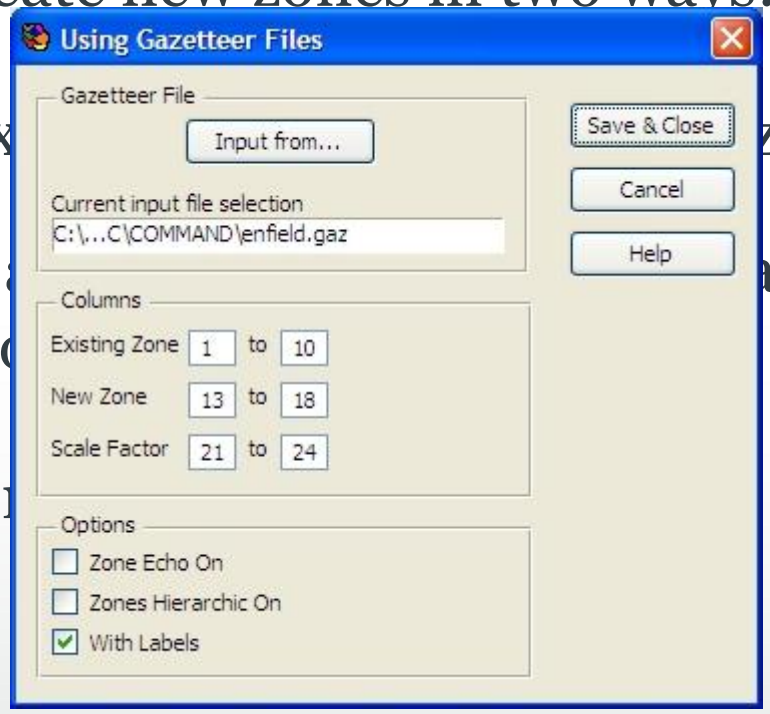
Data aggregation/rezoning: 2

- SASPAC can create new zones in two ways:

  1. Combining ex zones

  2. Reference to allocations of existing zones t

- In both cases, a



Data aggregation/rezoning: 3

- New zones may be created by:

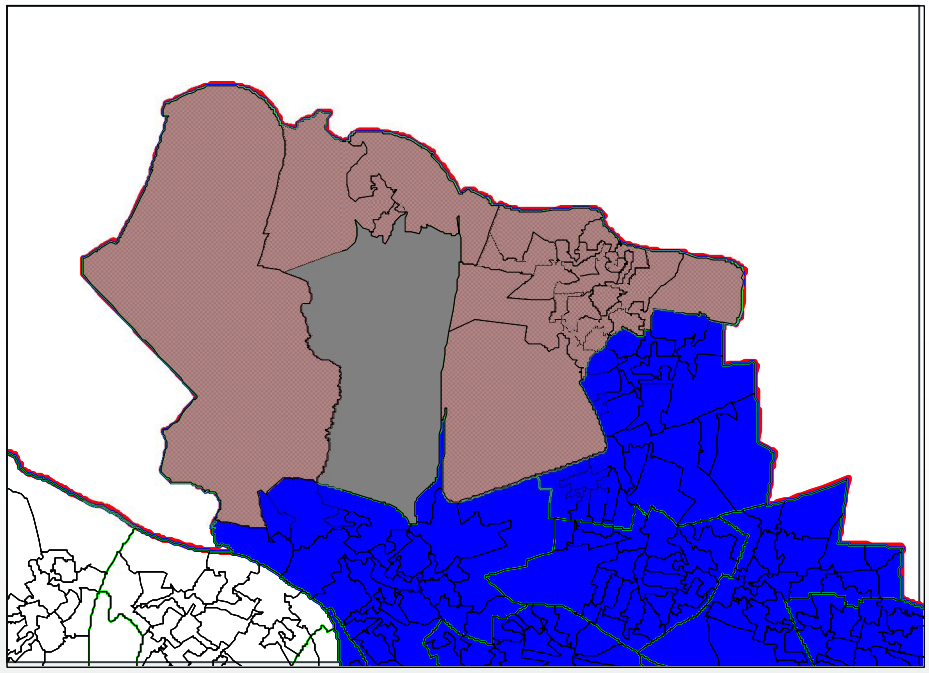
  1. Adding existing zones together

  2. Subtracting an area from another area of higher geographic level (e.g. ward from a LA)

  3. Use scale factors to allocate proportions of existing areas to new zones

Data aggregation/rezoning: 4

- For example:



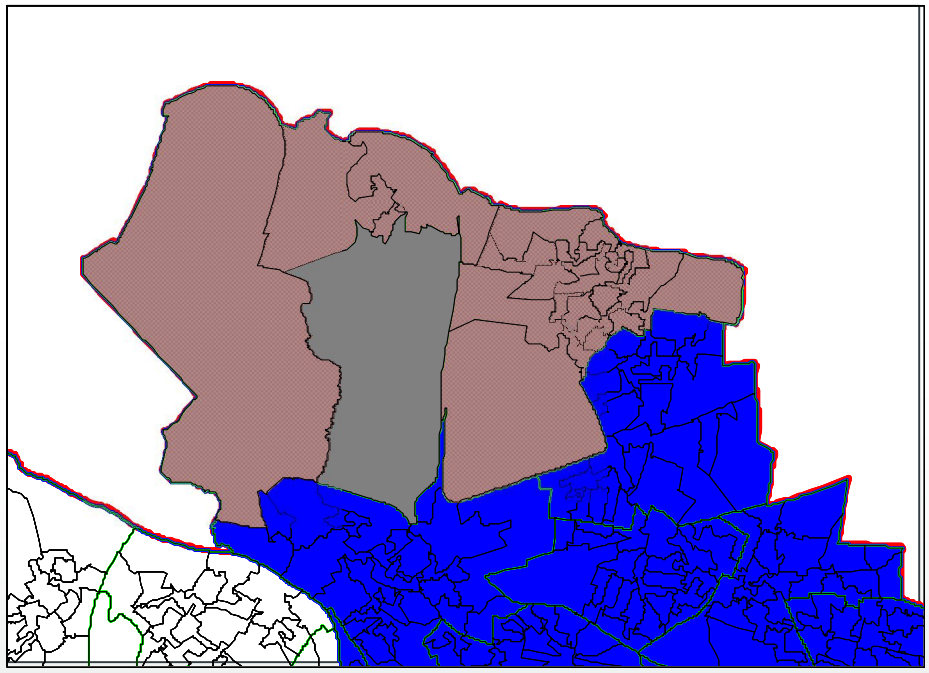
Spelthorne (LA)

Stanwell North (ward)

0018 (Output Area)

Police beat 1

Data aggregation/rezoning: 4

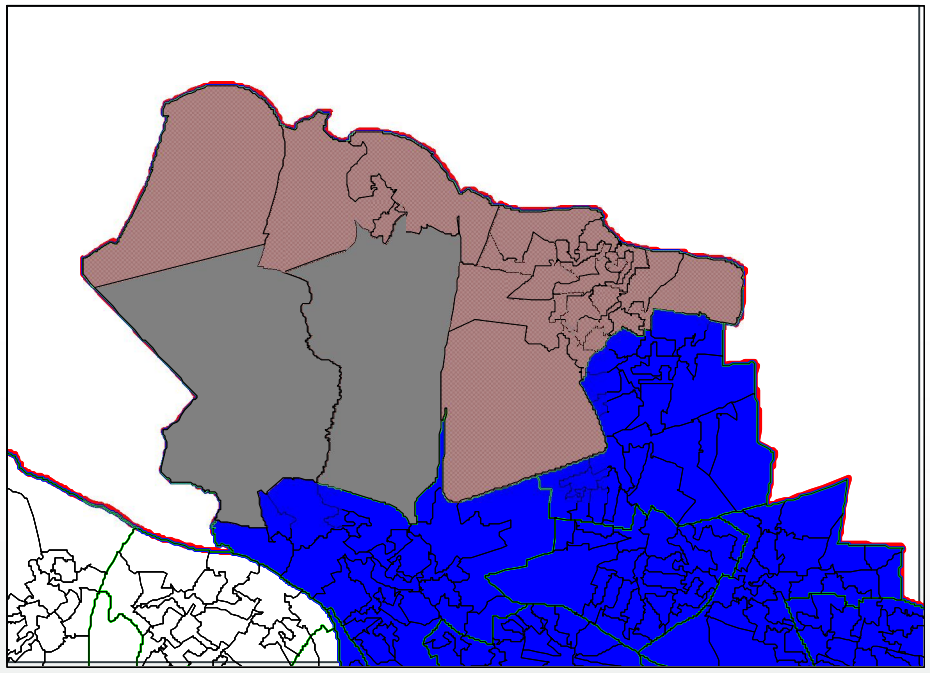


Police beat 1:

Stanwell North 43UHGD (without 0018)

….and 43UHGB0013

Data aggregation/rezoning: 4



Police beat 1a is a subset of beat 1, and consists of:

Stanwell North 43UHGD (without 0018)

….and ⅓ of 43UHGB0013.

Worked example: 2

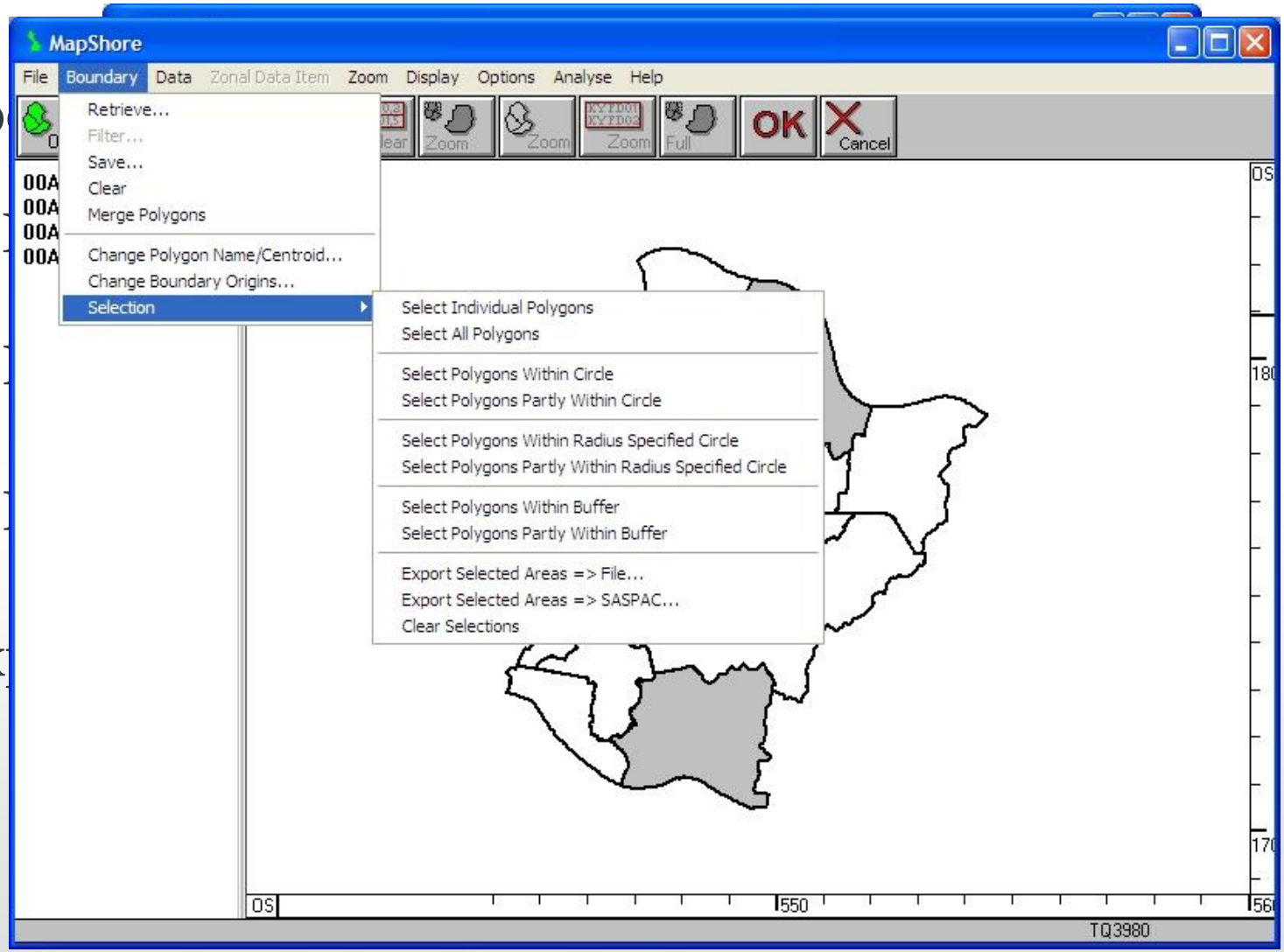
- Create a new zone that comprises of all the 2001 output areas that fall within 0.5km of Paddington rail station. Print the total resident population, males and females that fall within this new zone.

GIS – MapShore: 1

- SASPAC's integrated GIS
- native format is .bdy (import .shp/.MIF)
- 2001 Census Output Area boundaries available from helpdesk
- Other Census boundaries (ward, LAs) also available from Pebbleshore via the helpdesk

GIS – MapShore: 2

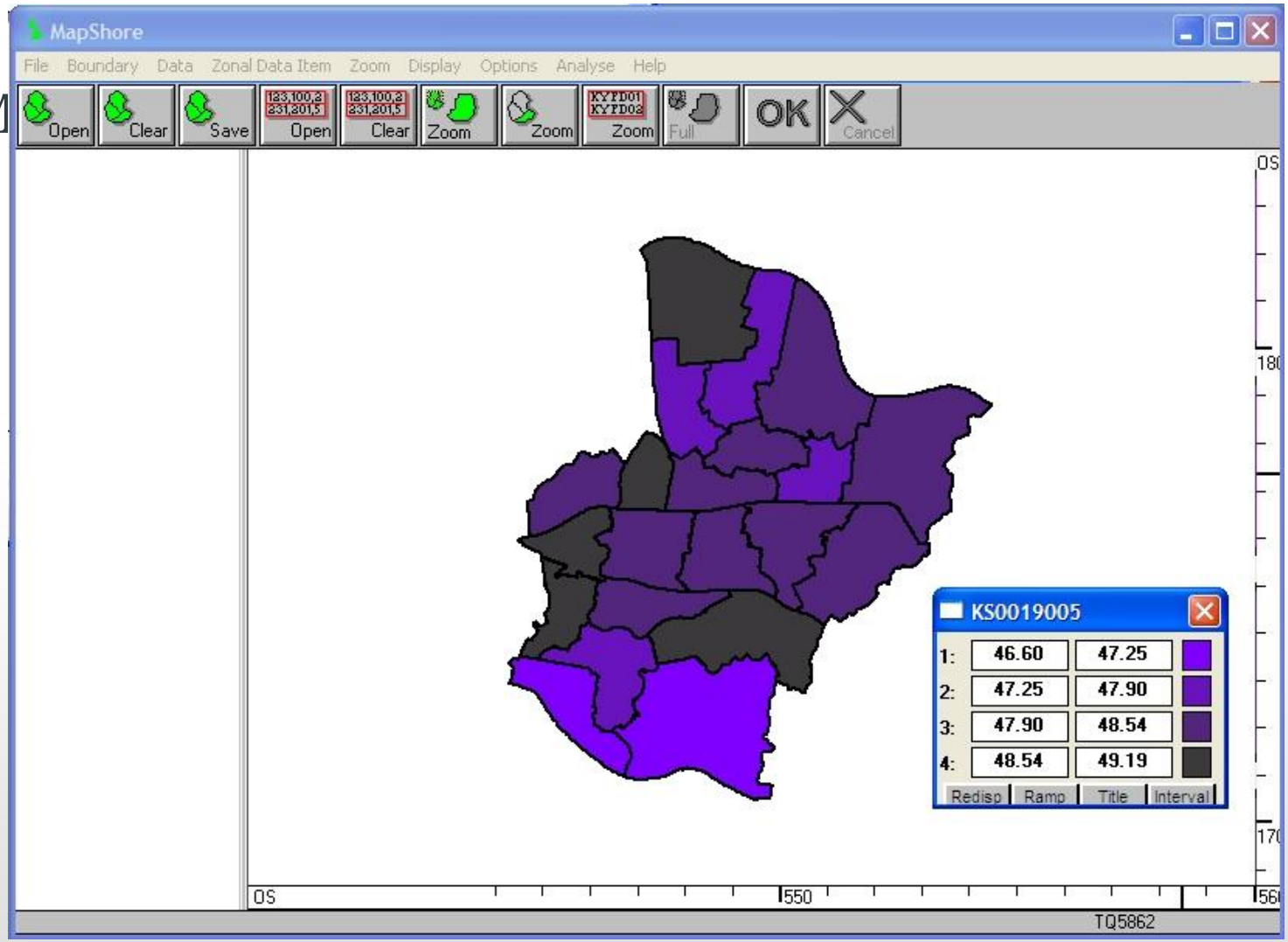
- To
  1.
  2.
  3.
- Ex



GIS – MapShore: 3

- M

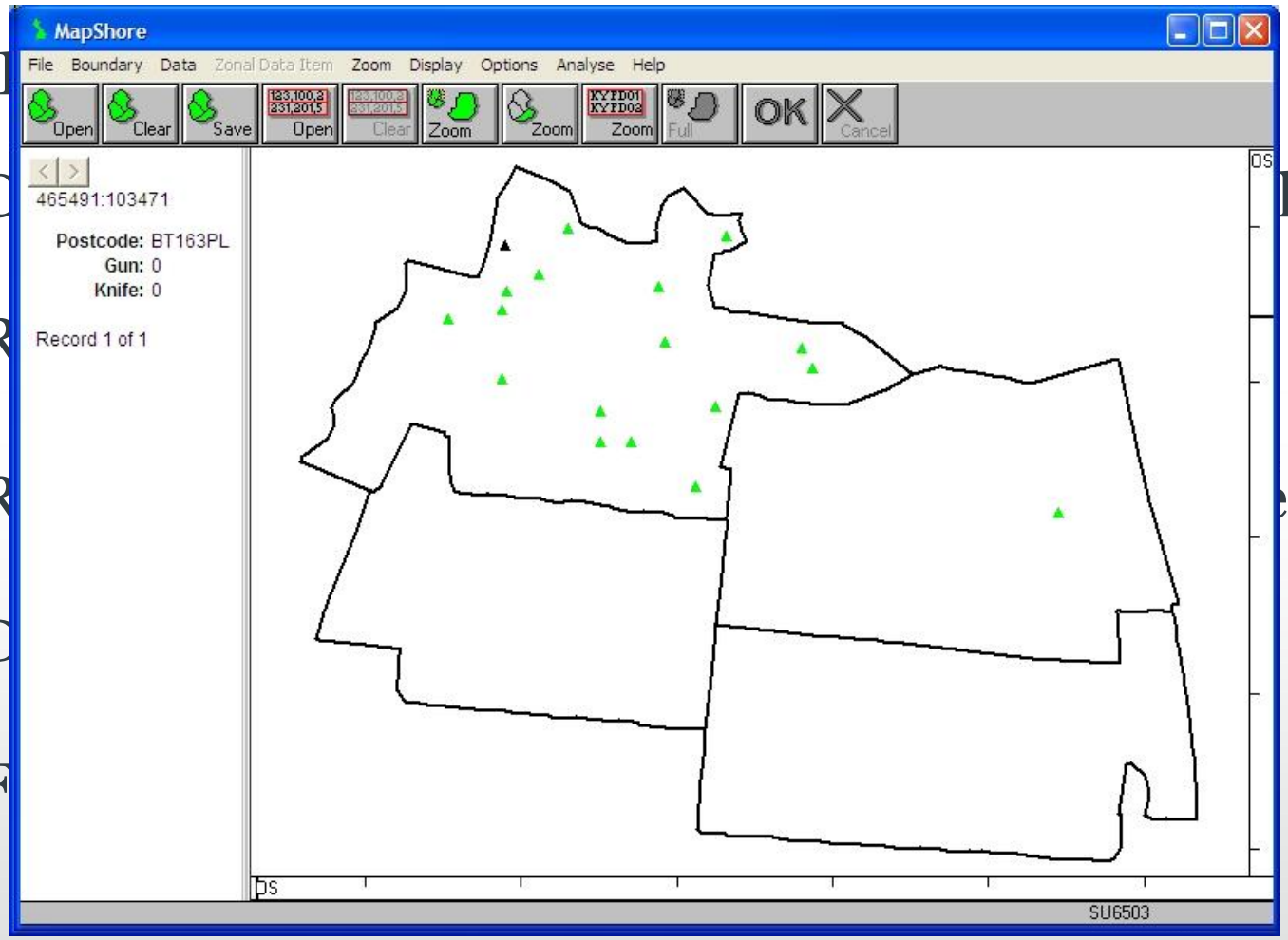
1.
2.
3.
4.
5.



GIS – MapShore: 4

- Map

1. C hotos
2. R
3. R e
4. D
5. F



Incidents of knife/gun crime

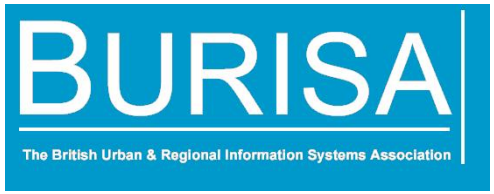
Worked example: 3

- Export and map a CSV file representing the percentage of residents who recorded their ethnic group as Bangladeshi in the 2001 Census for all the wards in the City of Westminster.

More help…

- www.saspac.org (news, examples, training manual, slides, videos…etc)
- Email support: info@saspac.org
- Phone support: 020 7983 4348
- Demos, training at City Hall/on location

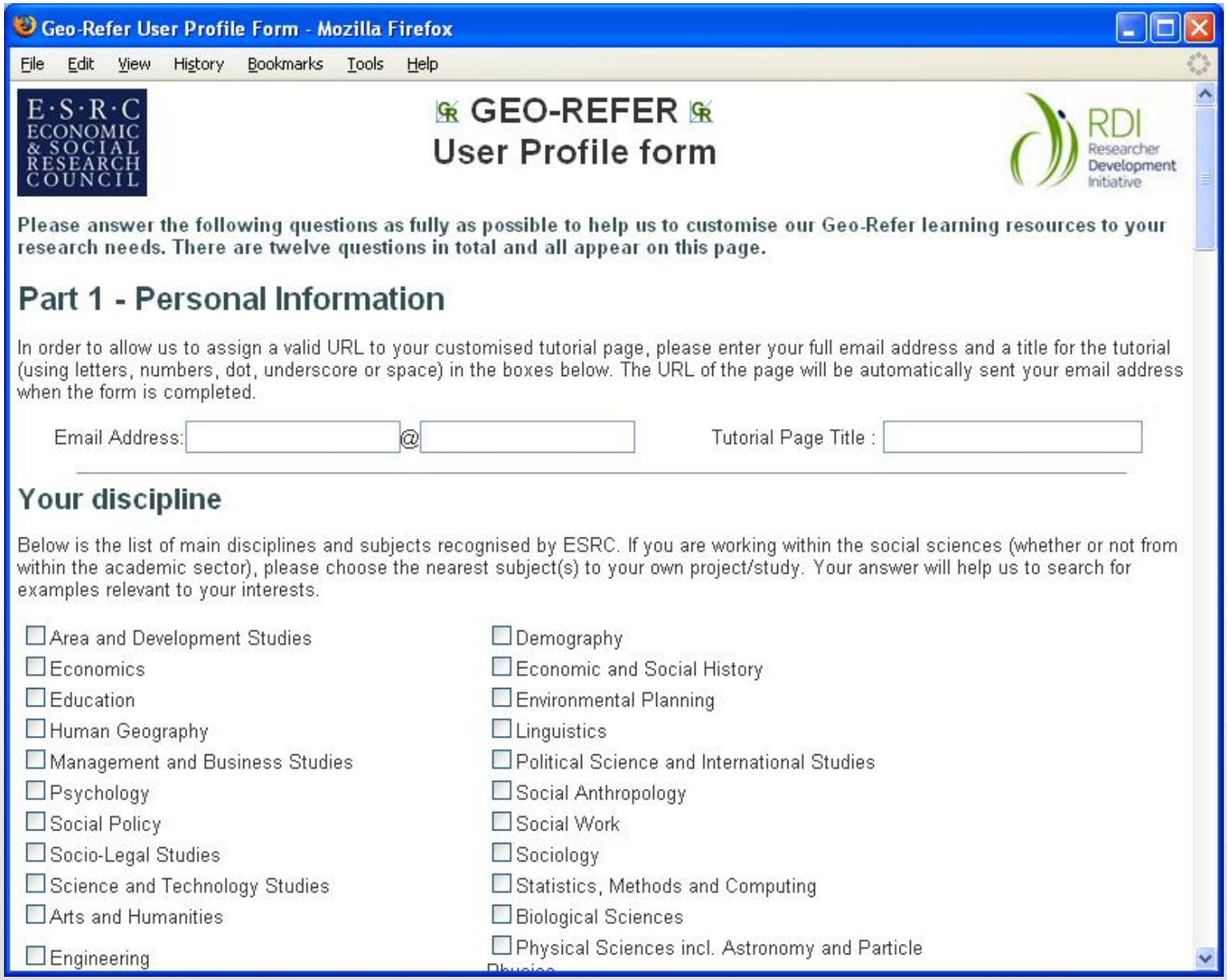
Briefing on afternoon practical sessions



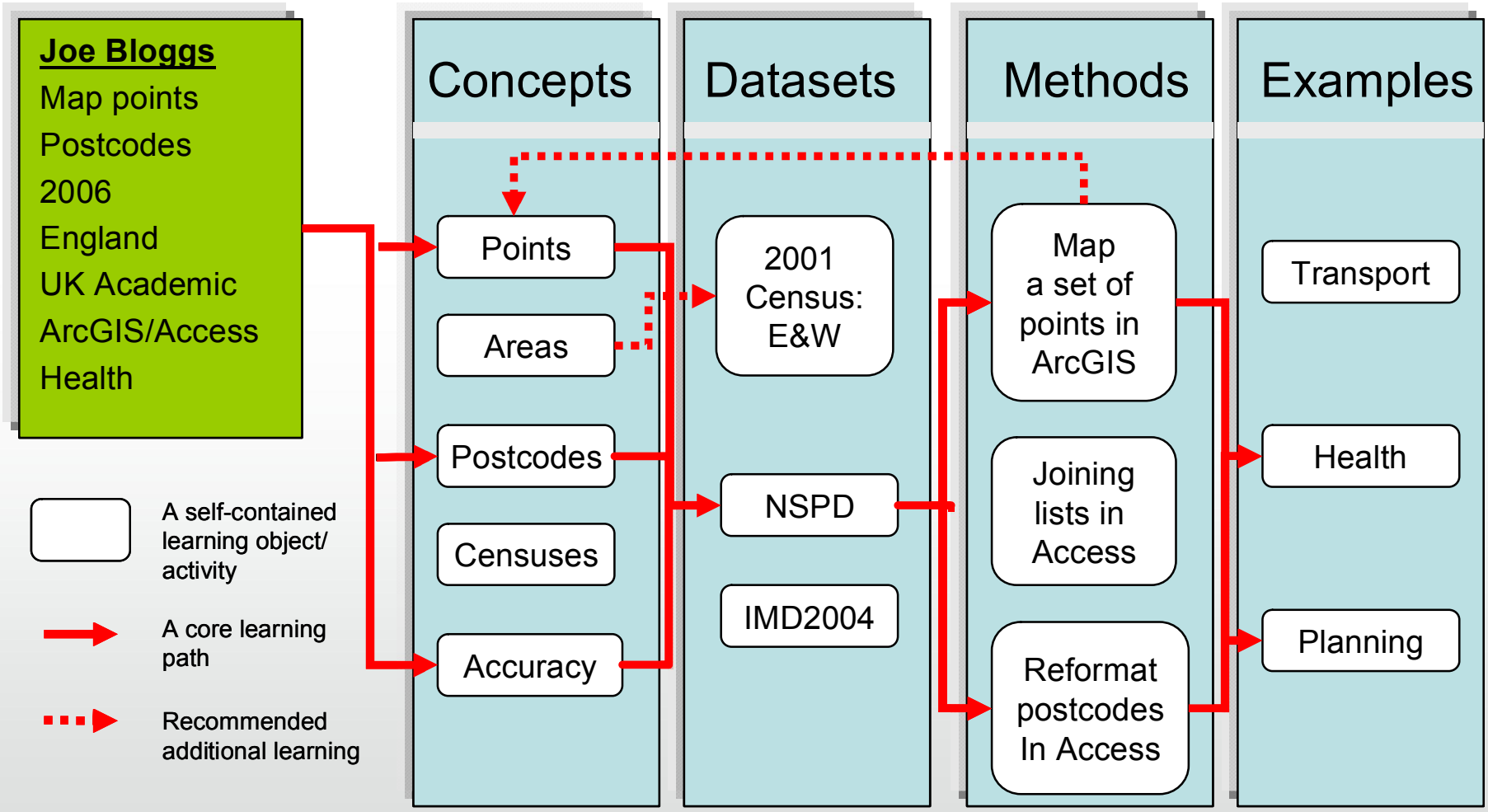
Using the Geo-Refer resources

- Complete the user profile form
  - Specify own profile and Geo-Refer will best-match learning materials
- Browse the resources
  - Extensive list of concepts, methods, datasets and examples
- URL - http://www.geog.soton.ac.uk/geo-refer/workshop5.html

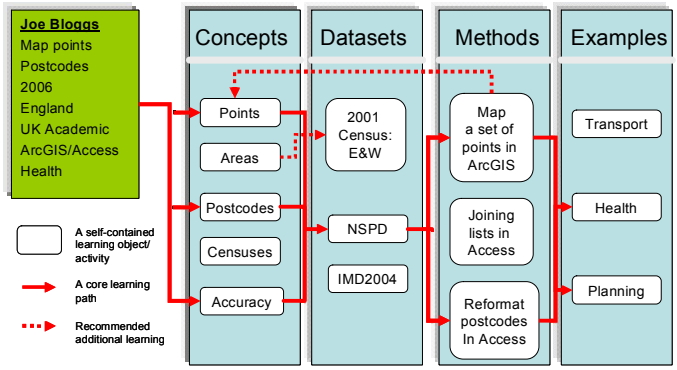
Geo-Refer user profile form



Customised set of learning resources



Customised online tutorial



Plan learning activities

- Think about what you've just heard
- Think about your own requirements
- Browse the Geo-Refer learning resources
- Identify what you'd like to work on (individually or in groups) and let us know

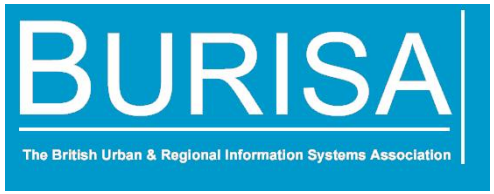
Some ideas

- Link performance table of the primary schools in a London borough to grid references using Microsoft Access
- Download ID2007 data from SASPAC and/or Neighbourhood Statistics service
- Retrieve Census and output area boundaries from SASPAC
- Map schools and deprivation information using MapShore
- Or use own datasets…

Lunch!!!

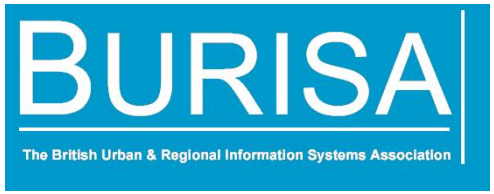

E·S·R·C
ECONOMIC
& SOCIAL
RESEARCH
COUNCIL


RDI
Researcher
Development
Initiative

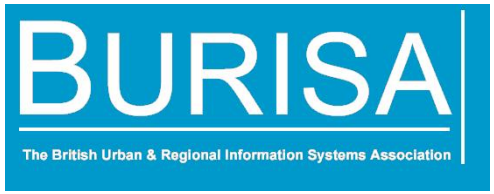

BURISA
The British Urban & Regional Information Systems Association


Census
Censas
Sensas
Sessas
Sassas
Saspas
Saspac

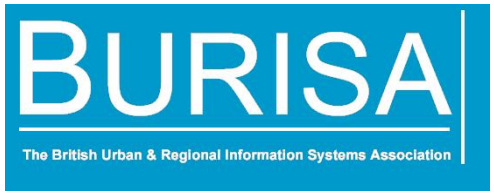
Round-up and feedback



Workshop evaluation form



Really useful tools/datasets



National Statistics Postcode Directory

Postcode

Ordnance Survey Grid Reference

Dates of introduction and termination (if applicable)

Delivery point count

2001 Census geography codes

1991 Census geography codes

Numerous other codes, esp. changing health geographies

Numerous derived lookup tables

AKA All Fields Postcode Directory, Central Postcode Directory, Postzon file…

Used in GeoConvert

Downloadable from UKBORDERS; registration required

http://ukborders.census.ac.uk

UKBORDERS Digital Boundary Data

Coordinates of area boundaries downloadable in various GIS and mapping formats

Primarily census-derived, covering whole UK

Census, administrative, health and electoral geographies

Includes some historical (pre-1971 boundary sets e.g. 1951 local government

Need to understand which zones are needed

Need suitable software to load the boundaries

Large data volumes and further manipulation often required

No attribute data

Downloadable from UKBORDERS; registration required

http://ukborders.census.ac.uk

Neighbourhood Statistics Services

Wide range of easily downloadable social data

Includes basic 2001 census datasets and administrative data derived from government departments

Searchable by many different georeferences

Some online mapping and analysis

Simple – general audience

Only covers most basic datasets

No lookup tables or boundary data downloadable, just statistics for areas

Data formatting can be frustrating for serious user

Freely accessible from ONS, GROS, NISRA websites; registration increases functionality

http://www.neighbourhood.statistics.gov.uk

ONS Beginners Guide to UK Geography

Online explanatory material about contemporary UK boundary systems

Includes census, postal, administrative, electoral, health and other geographies

Useful links to other resources

Very clear – written for general audience

Good reference for students!

Does not cover historical datasets: essentially describes the post-2001 census situation

No downloadable data, just explanatory guidance

Freely accessible from ONS website

http://www.statistics.gov.uk/geography/beginners_guide.asp

Other useful sites…

Royal Mail Postcode Finder and Address Finder http://www.royalmail.co.uk	> Registration required and limited number of uses
CASWEB http://census.ac.uk/casweb	> Downloadable census statistics (1971-2001) through an online interface. Boundary data only for 1991. Registration required.
Digimap – Ordnance Survey mapping, inc. historical www.edina.ac.uk/digimap	> JISC-funded subscription service at Edina
Google Earth and Microsoft Live Search Maps www.earth.google.com www.maps.live.com	> Street mapping and aerial photography (and many others)