# Linking and mapping geographically referenced data

**David Martin** 

NCRM Autumn School, 13 November 2008





#### Overview of session

- Some practicalities
- The Geo-Refer projects
- Geographical referencing concepts
- Georeferencing methods: linking data
- Georeferencing methods: mapping data
- Getting more help: how to use the Geo-Refer online resources

#### Purpose of session

- Challenge participants to consider whether there might be geographical aspects to their own work
- Overview issues involved in geographical referencing of social, economic and environmental datasets
- Some pointers to useful methods and techniques
- Key points in relation to a few essential datasets
- Information about online resources that will help you go further

#### Some practicalities

- Presentation based on one more often used in association with a practical hands-on session
- Here, restricted to demonstration and discussion
- This is an informal session it will work best if interactive
- There is no such thing as a silly question!
- Please interrupt

## Why would I want to know about georeferencing and mapping?

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

#### Audience participation...

What data do (or might) you use that are (or could be) geographically referenced?





#### The Geo-Refer Projects

- ESRC Researcher Development Initiative
- 2 Geo-Refer projects
- With colleagues Samantha Cockings and Samuel Leung
- Focus on training resources in geographical referencing issues
- Particularly aimed at users who are not geographers/ geographical information specialists
- Training workshops and online resources

#### Georeferencing examples

- Link survey results to census data
- Associate postcoded patient records to indices of deprivation
- Produce a map of local authorities showing response rates
- Allocate questionnaire responses to urban/rural neighbourhood types
- Identify which service delivery locations fall within different policy areas

## Geo-referencing concepts and methods





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



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



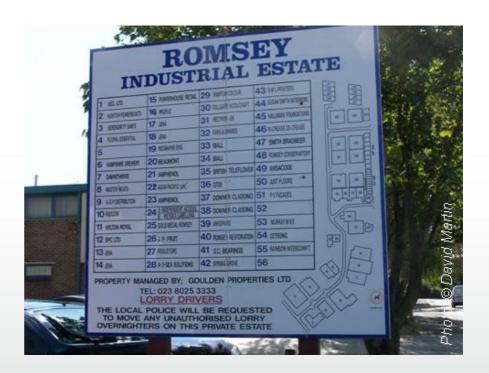
- Phenomena that are linear, describing routes or flows
- e.g. bus route, telephone conversation, social network link



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



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



### Geographical object types



#### 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

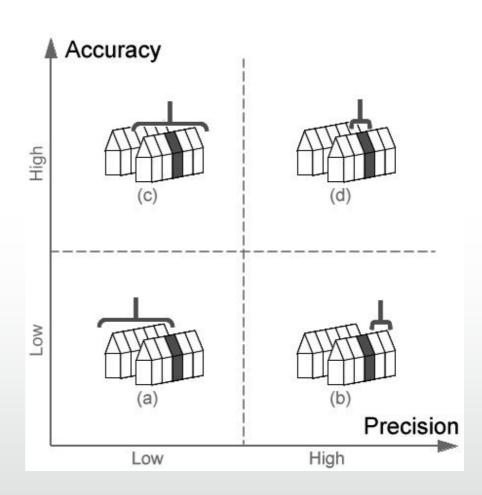
- Direct georeferencing: explicit coordinate system
  - Lat/Long, Ordnance Survey grid reference
- Indirect georeferencing: any type of area code or name relating to a known location
  - Postcode, Zip code, County, Census output area/enumeration district/tract; local government district; health authority,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

#### Where do coordinates come from?

- Surveys for mapping
- Remote sensing
- Digitising paper source documents
- Direct capture using Global Positioning System (GPS) receiver
- In the social sciences, geographical coordinates are generally secondary data, even when embedded within primary data!

### 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 at25b Highfield Lane
- Need to assess and adopt different strategies according to purpose

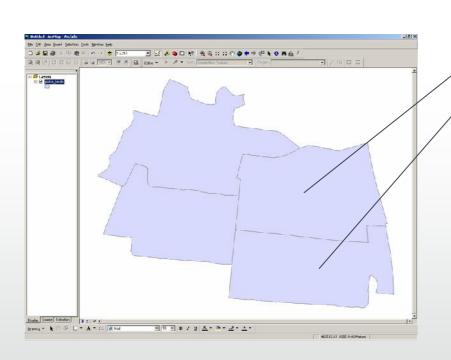


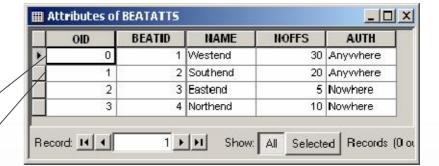
### Scale and projection

- 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: 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, so give consideration when considering research ethics

#### What about geographical information systems?

- Massive growth in use of GIS since early 1980s: complex software systems
- GIS growth promoting data standards, growth of geoinformation industry etc.
- GIS provide useful georeferencing tools and lots of other functions not needed here...
- ArcGIS and MapInfo (and others) widespread in UK universities
- This is not a GIS workshop!

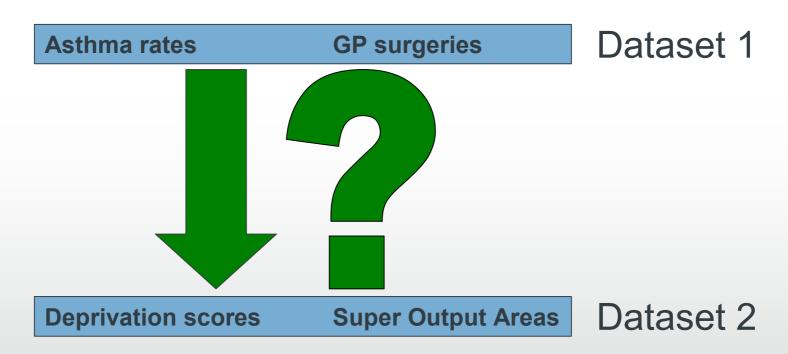
# Georeferencing methods: linking data





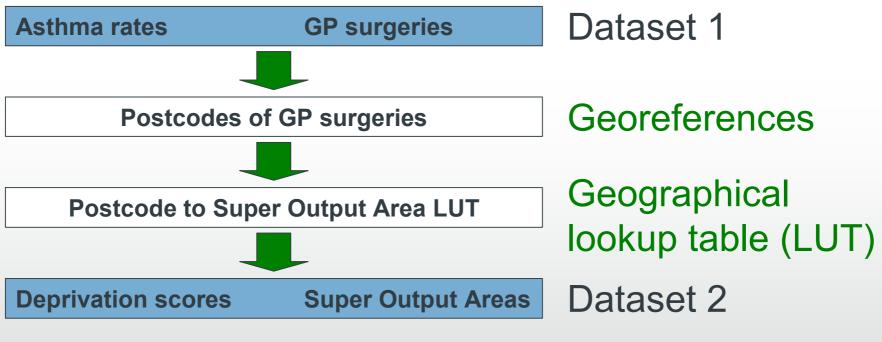
### Tabular data linkage example

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



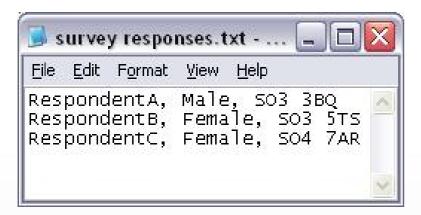
Assumes surgery is appropriate georeference for patient data!

#### Tabular linkage tools

- Access, SPSS etc. general purpose database or statistical packages for matching lists with common data items
  - Probably necessary for large or non-standard data linkage tasks
- GeoConvert online tool specific to recent UK postcodebased lookup
  - Very useful for a wide range of small-medium standard UK data linkages

#### Data linkage example: input tables

#### Coded survey responses

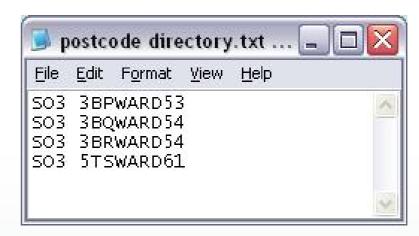


Respondent ID

Gender

Postcode

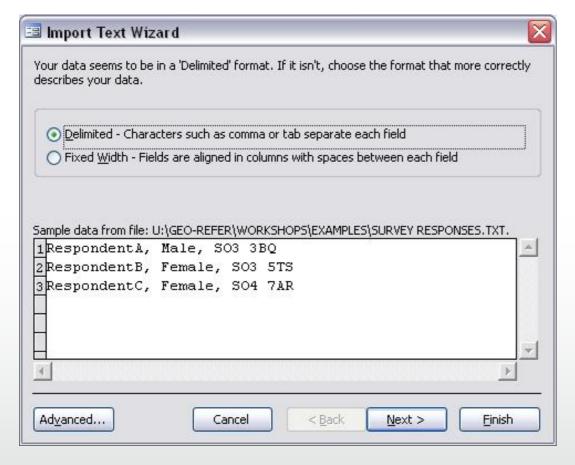
#### Postcode directory



Postcode

Ward code

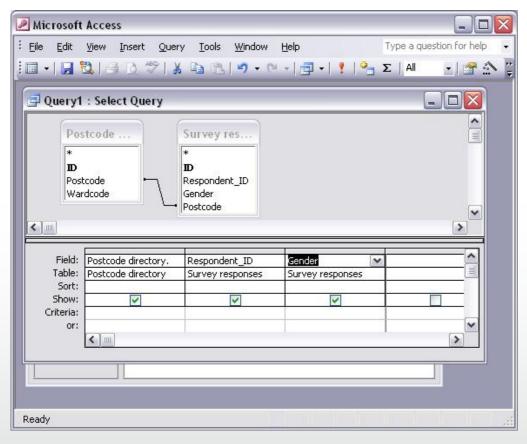
#### Import external data files



Get external data > import

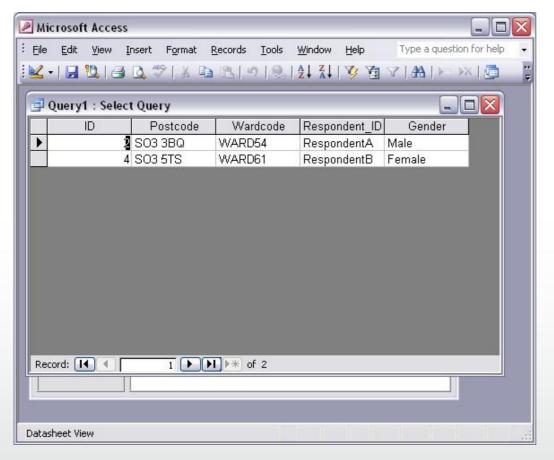
We need to describe the data structure (here comma separated)

#### Build new Access query



We now add both tables to a new query, link them by the postcode field and select the items required in output

#### Matched lists



The resulting query contains requested fields for all matching rows and can be exported

NB not all are matched!

#### 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

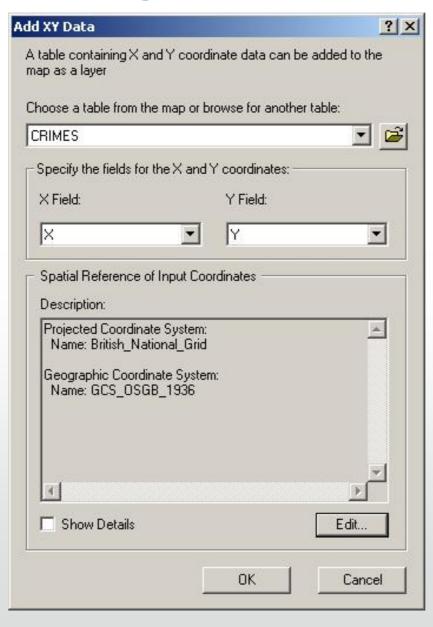




### Mapping

- 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

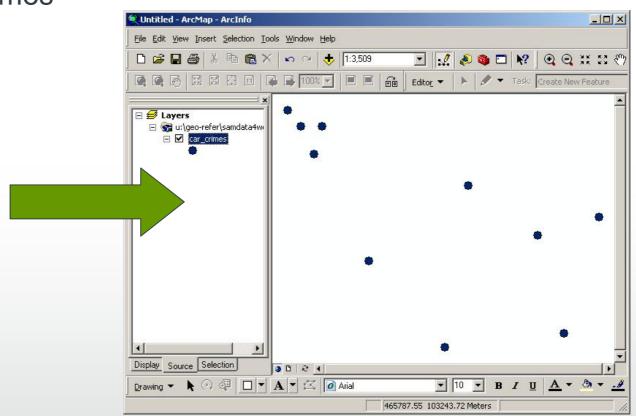
### Mapping a set of grid-referenced points



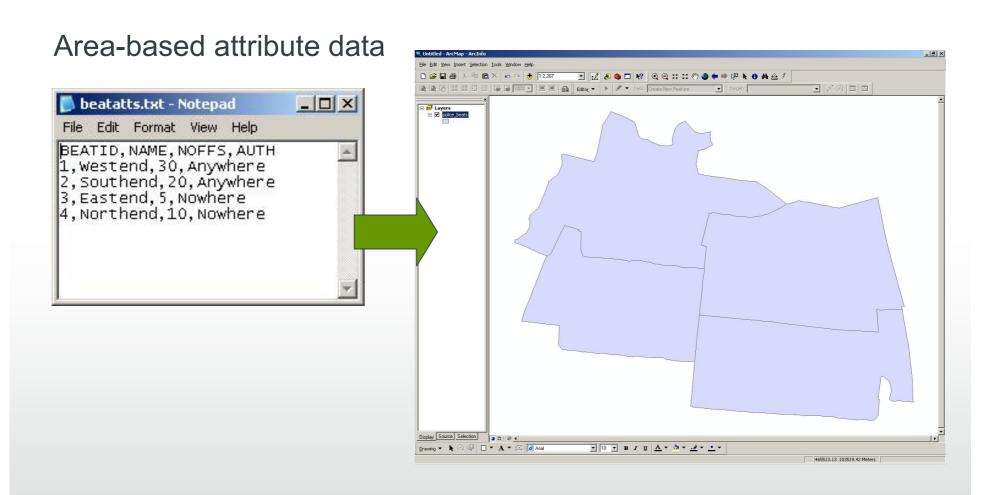
#### Mapping grid-referenced points

#### 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

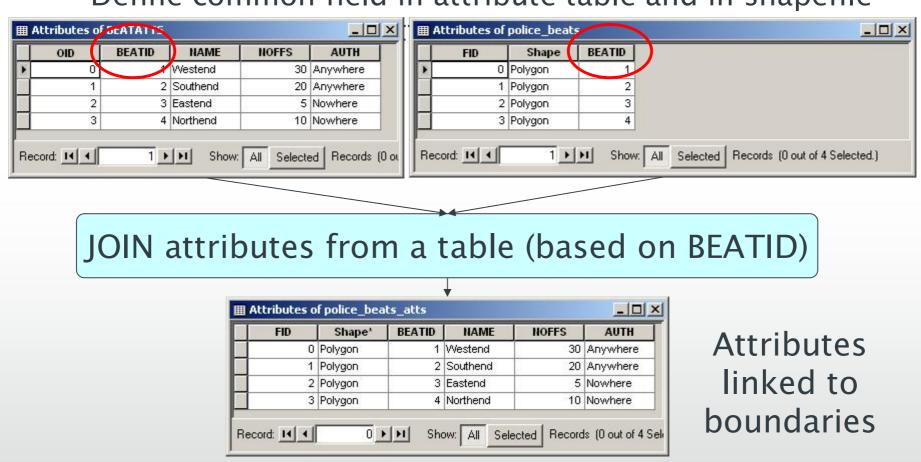


# Linking attribute data to boundary data



# Linking attribute data to boundary data

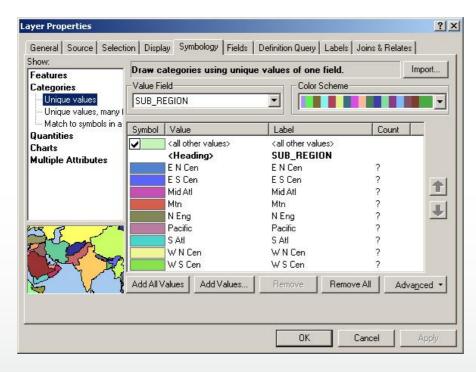
Define common field in attribute table and in shapefile



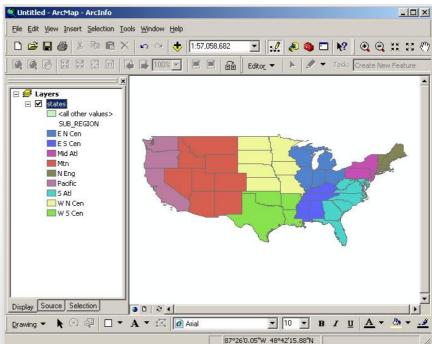
# Creating a choropleth/thematic map

- Add layer (shapefile)
- Select variable to map
- Select data type
- (Select classification scheme/class breaks)
- Select colour scheme

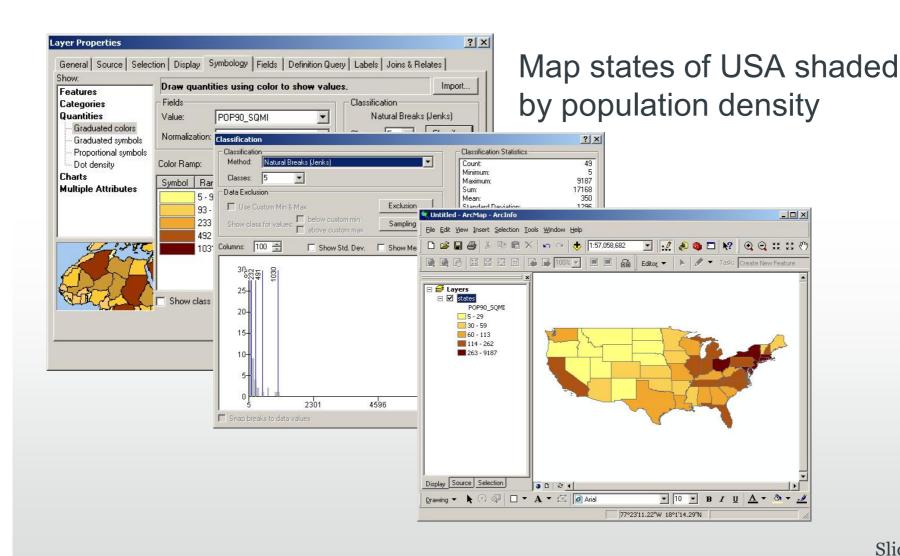
# Mapping area data: Categorical



# Map states of USA shaded by sub-region



### Mapping area data: Continuous

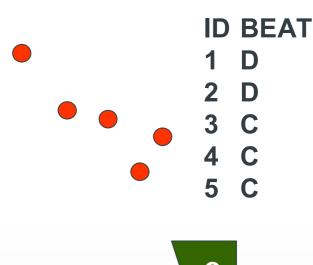


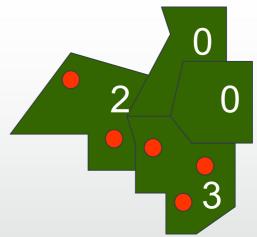
# Spatial data linkage

- 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





# 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

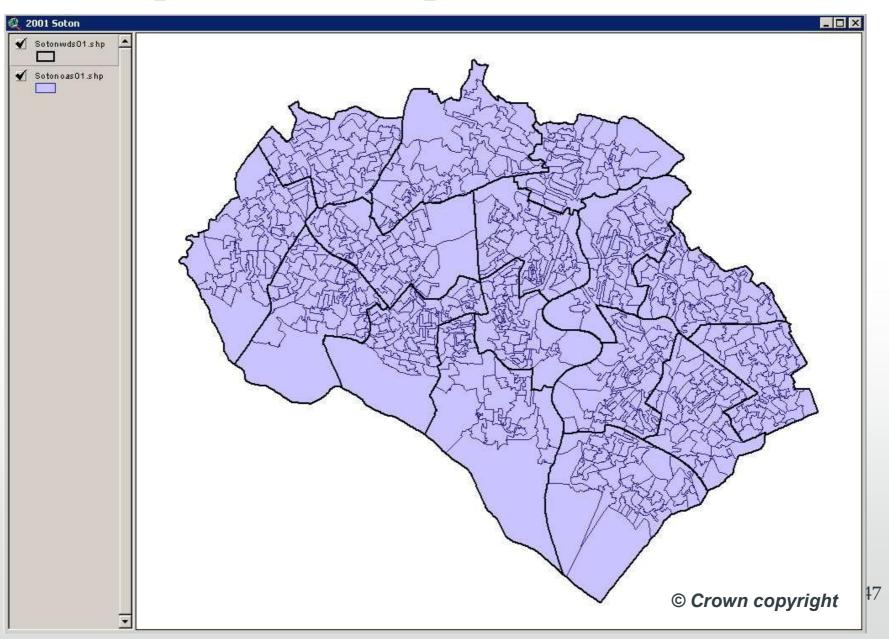




# Some very useful geographically referenced datasets

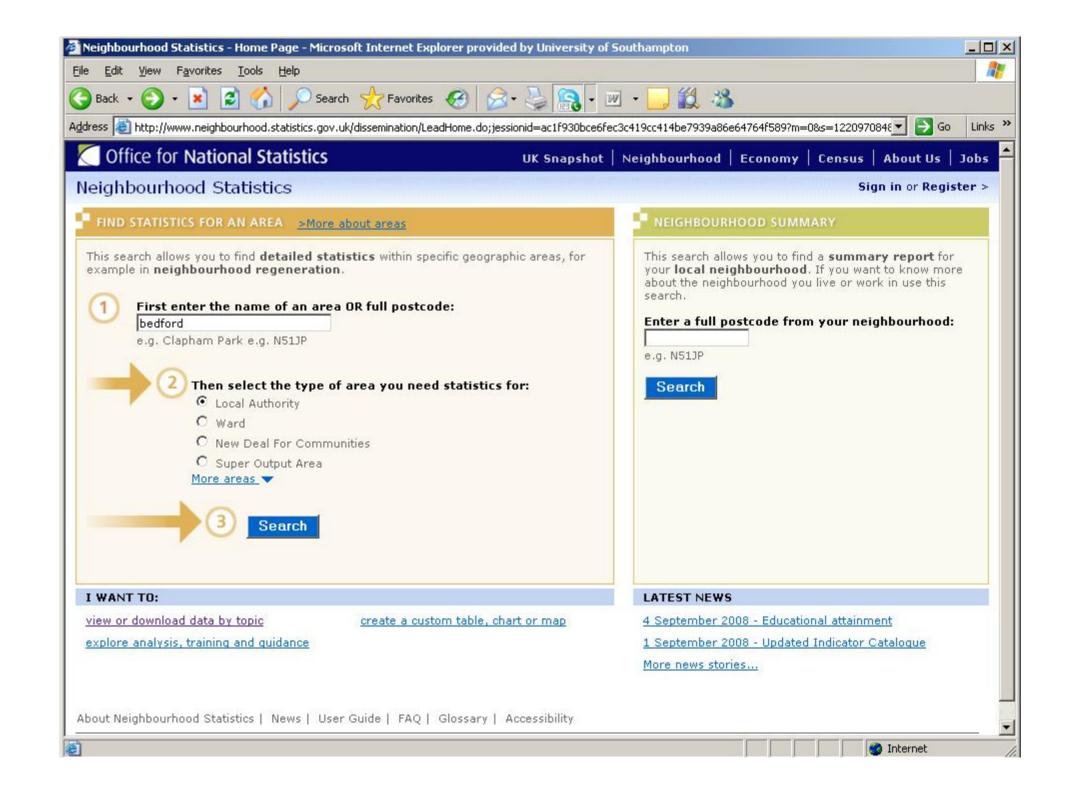
- Demographic data
  - Detailed populations: 2001 census (OA, ward, LAD)
  - Updated populations: mid-year estimates (LSOA)
  - Neighbourhood Statistics (OA/LSOA/MSOA)
- Indicators and classifications
  - Urban/rural codes, OA classifications (OA)
  - Indices of Deprivation 2007 (LSOA)
- Linkage datasets
  - National statistics postcode directory

### Southampton: 2001 output areas and wards



# Demographic data: census codes and boundaries

- Area hierarchy and codes
  - Output Areas (no names) 25JPUF0005
  - Ward UF
  - LADs (county/district 25JP or unitary authority 00JP)
- Digital boundaries
  - Available all levels from OA upwards
- OAs mostly built up from postcodes
- Census geography changes each decade (until 2011?)



# Linkage datasets: National Statistics Postcode Directory

- Full list of UK postcodes, updated quarterly
- Dates of introduction and termination
- Ordnance Survey grid references
- Multiple area codes, including
  - Census, neighbourhood statistics, local government, health, electoral

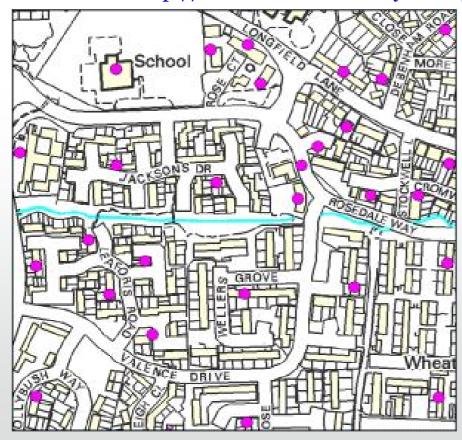
### Postcodes: coding structure

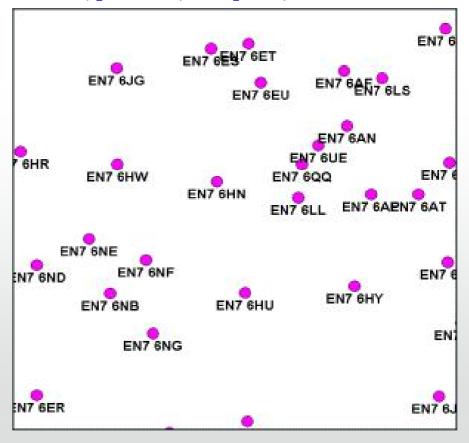
- Smallest standard geographical codes
- Integrated with addressing (listed in Royal Mail PAF)
- Area hierarchy and codes
  - Unit postcode (NOT areas) SO17 1BJ (~15 addresses)
  - Sector SO<sub>17</sub> 1
  - District SO17
  - Area SO
- Widespread recognition and use

# Postcodes: spatial referencing

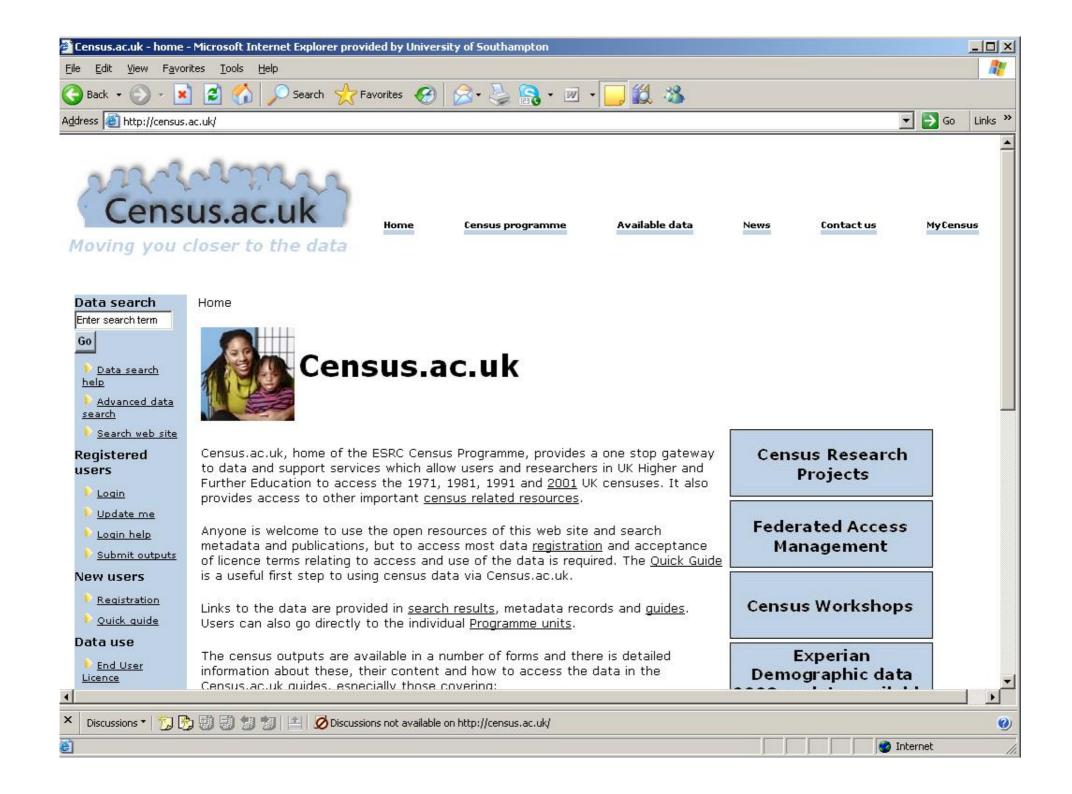
#### e.g. Ordnance Survey Code-Point

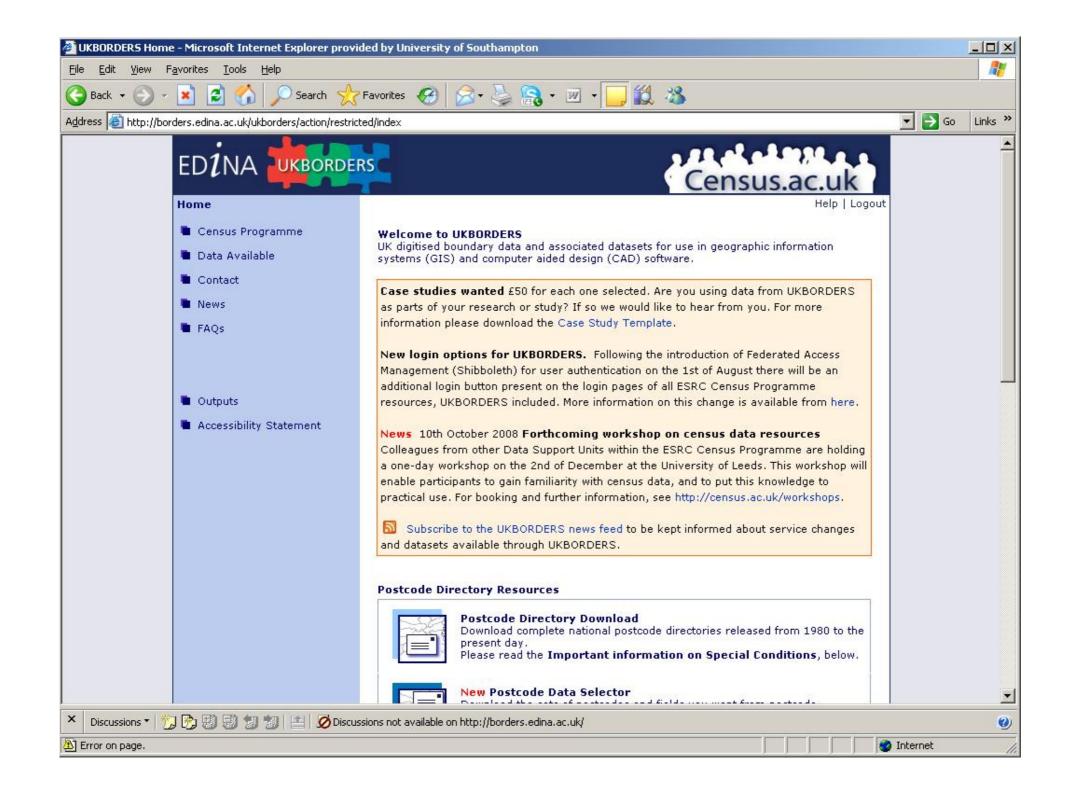
http://www.ordnancesurvey.co.uk/oswebsite/products/codepoint/

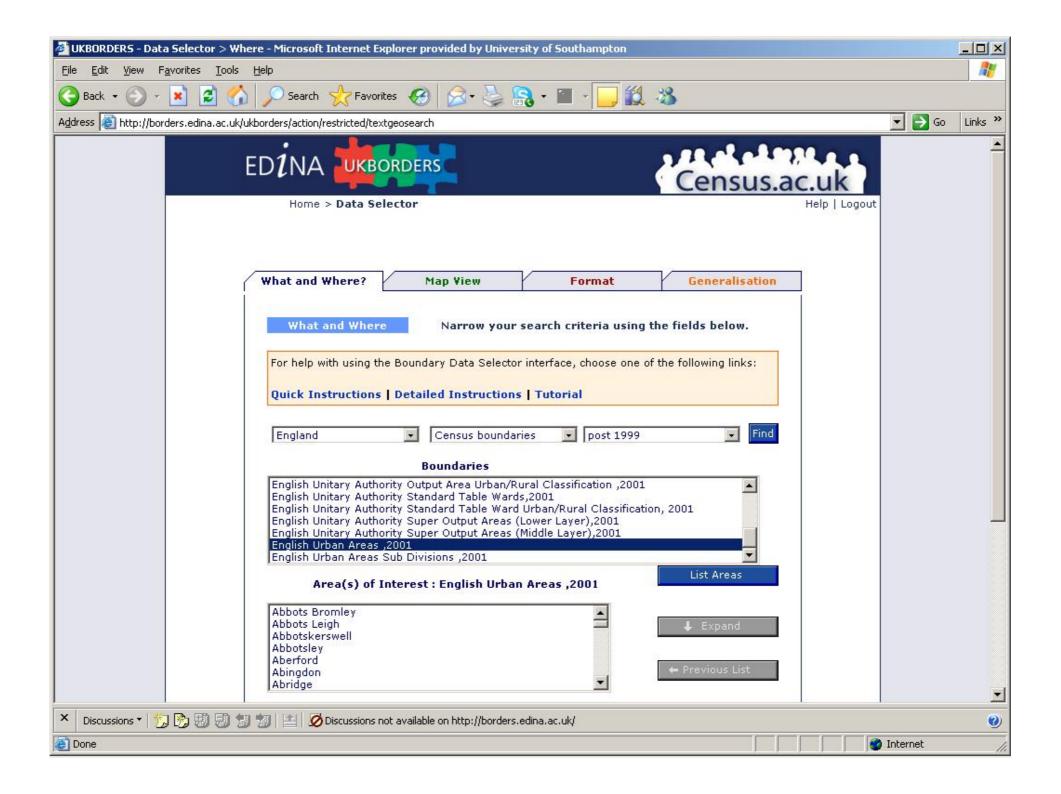


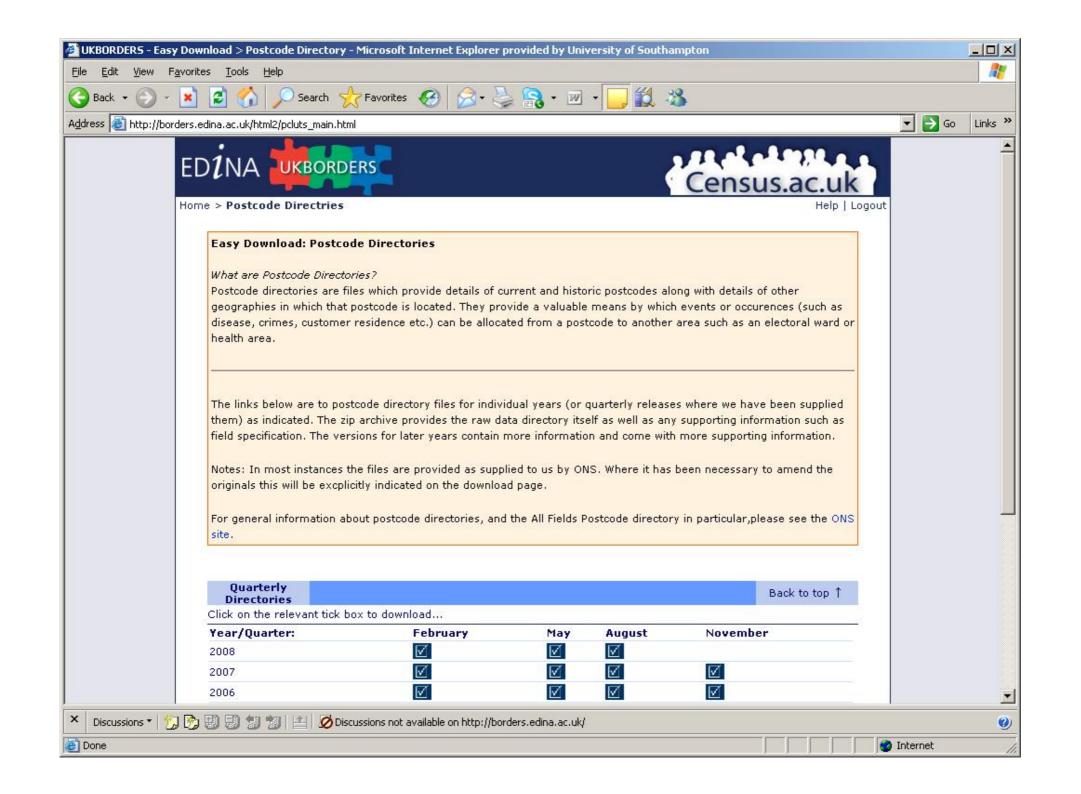


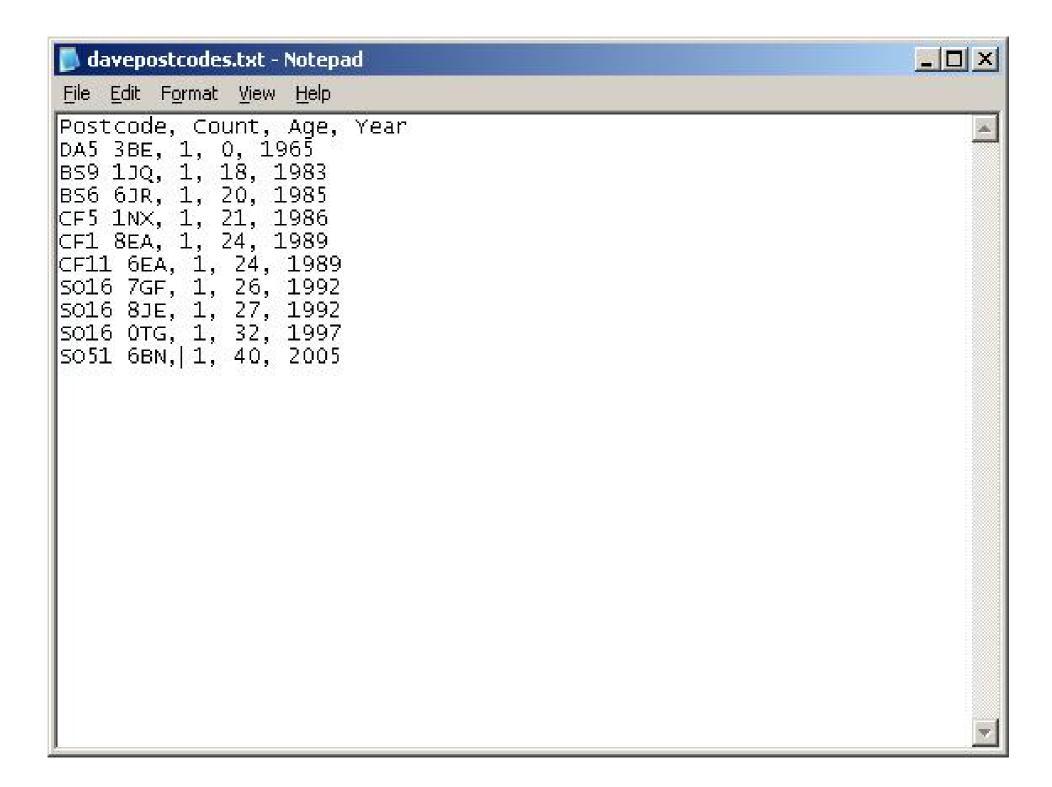
Demonstration of online services: Census.ac.uk, UKBorders, Geoconvert

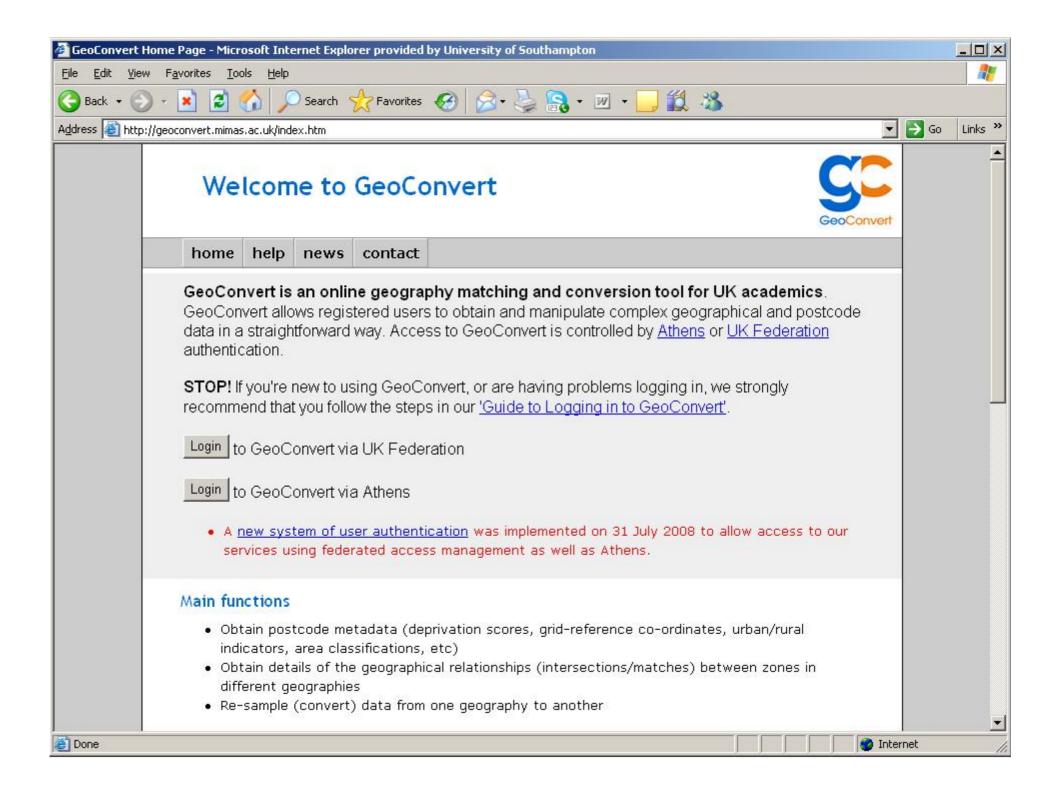


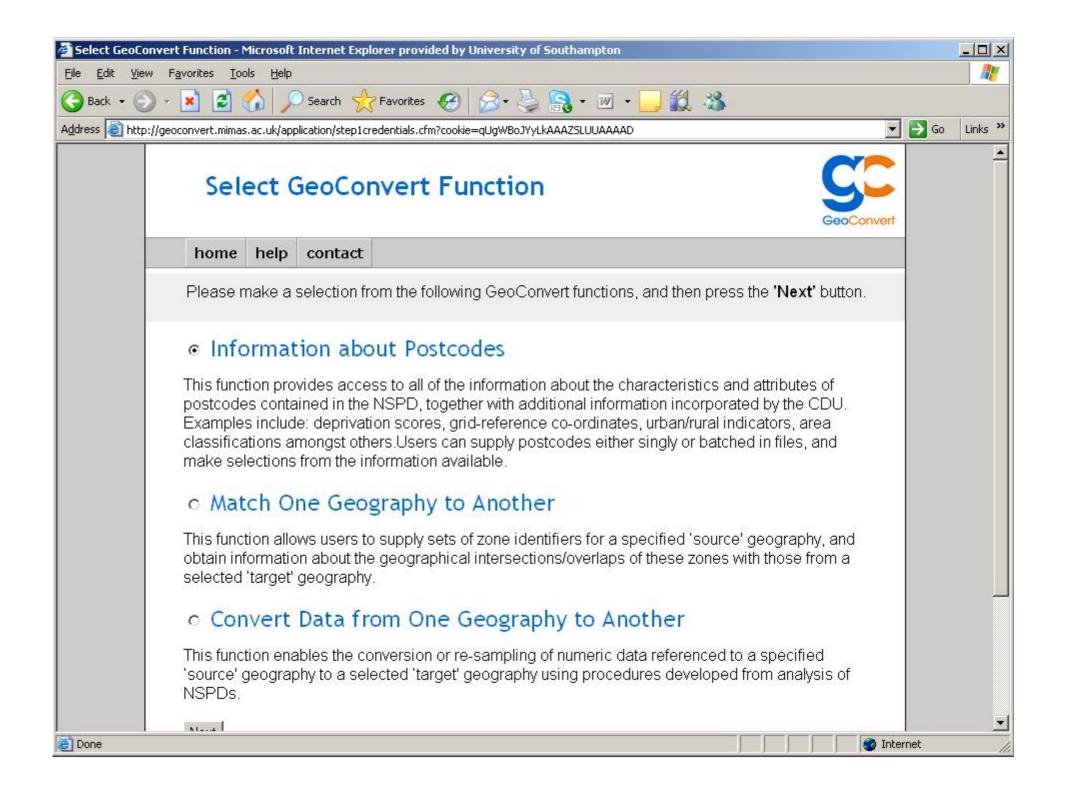


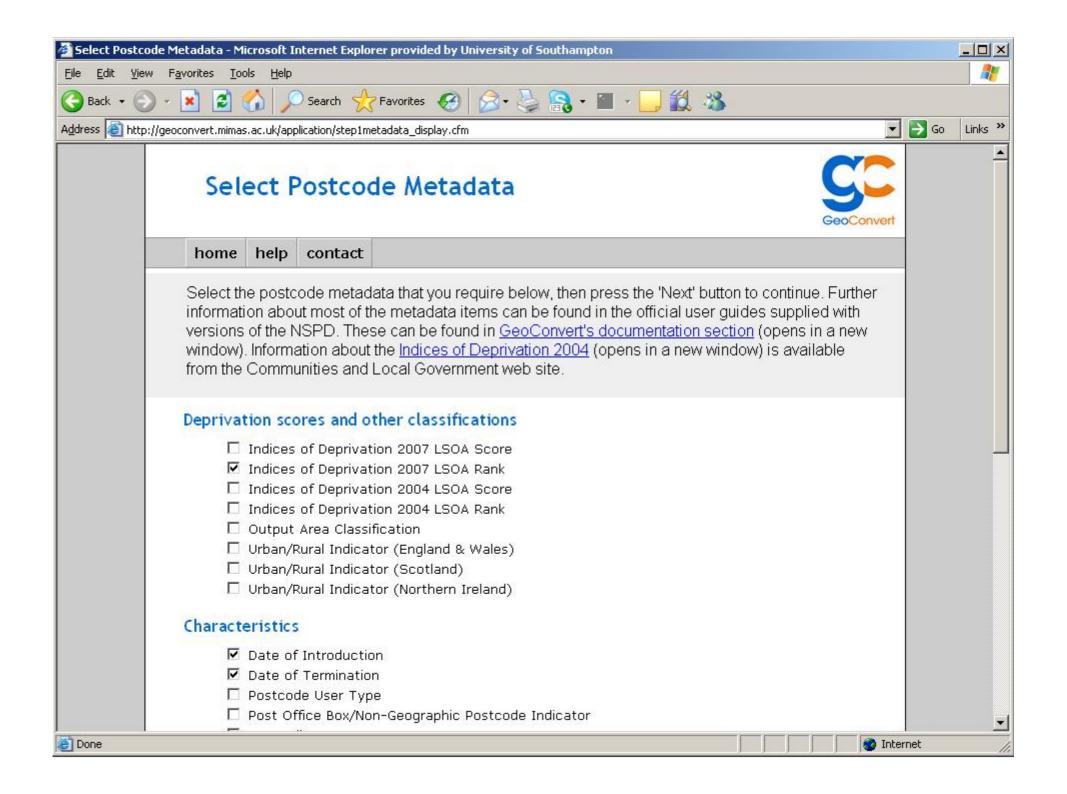


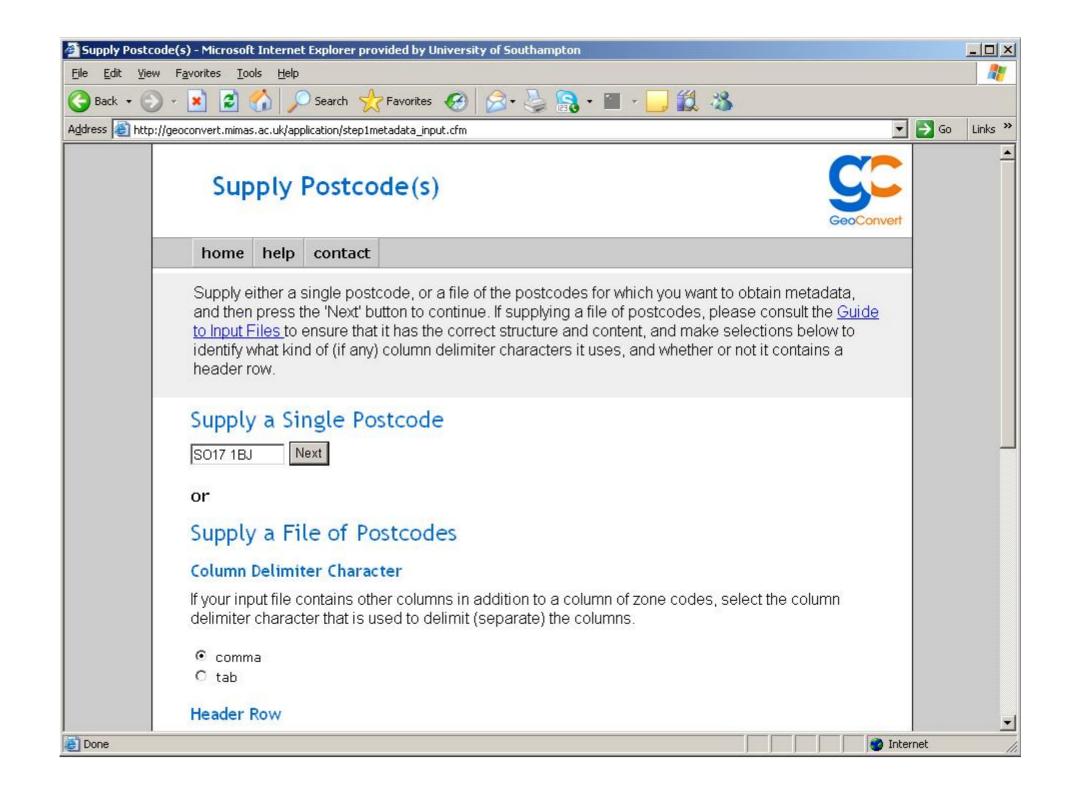


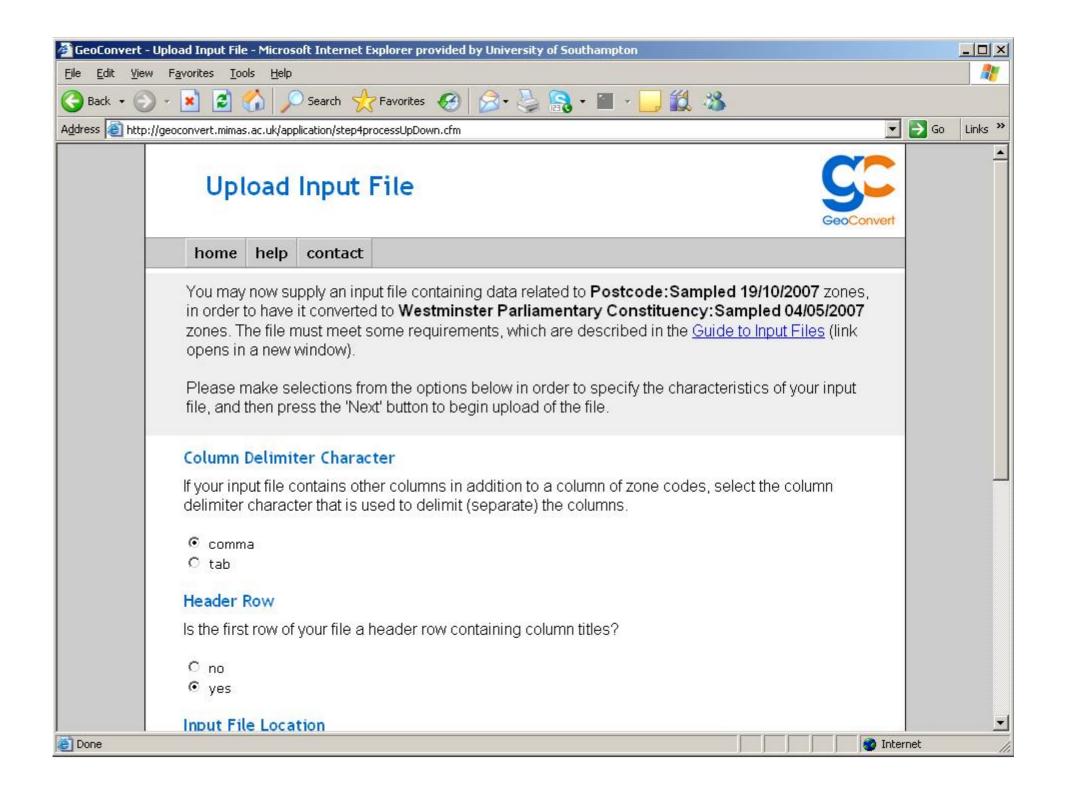


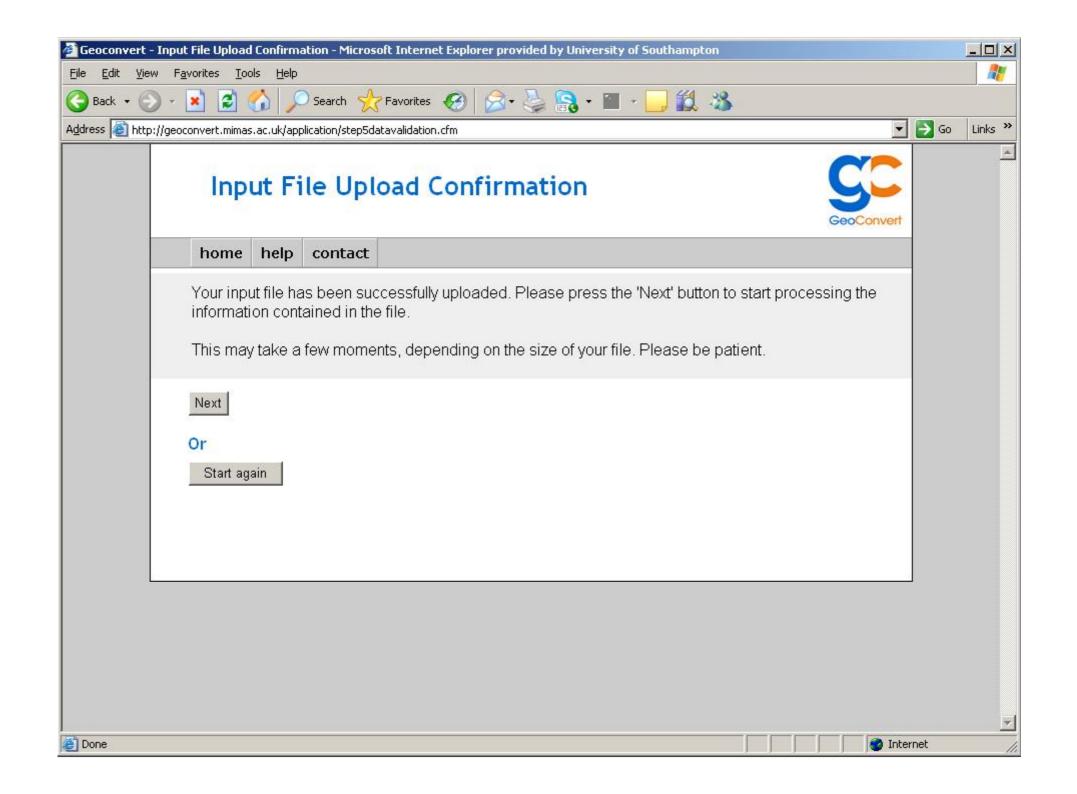


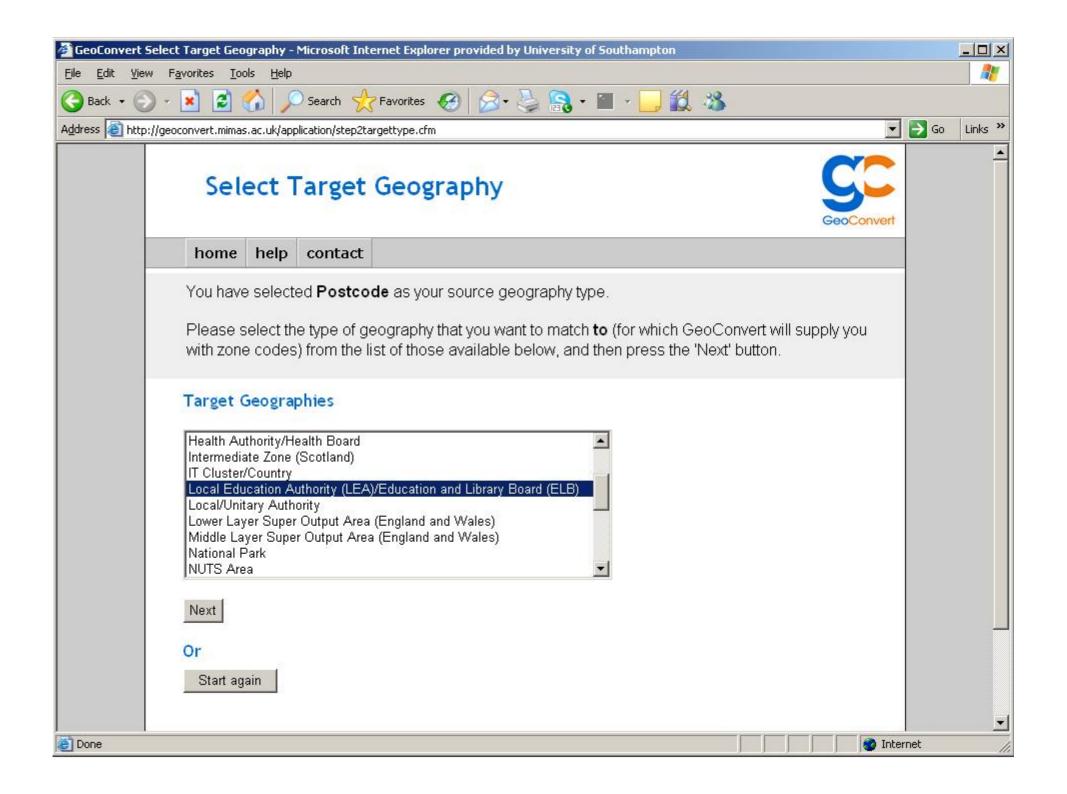


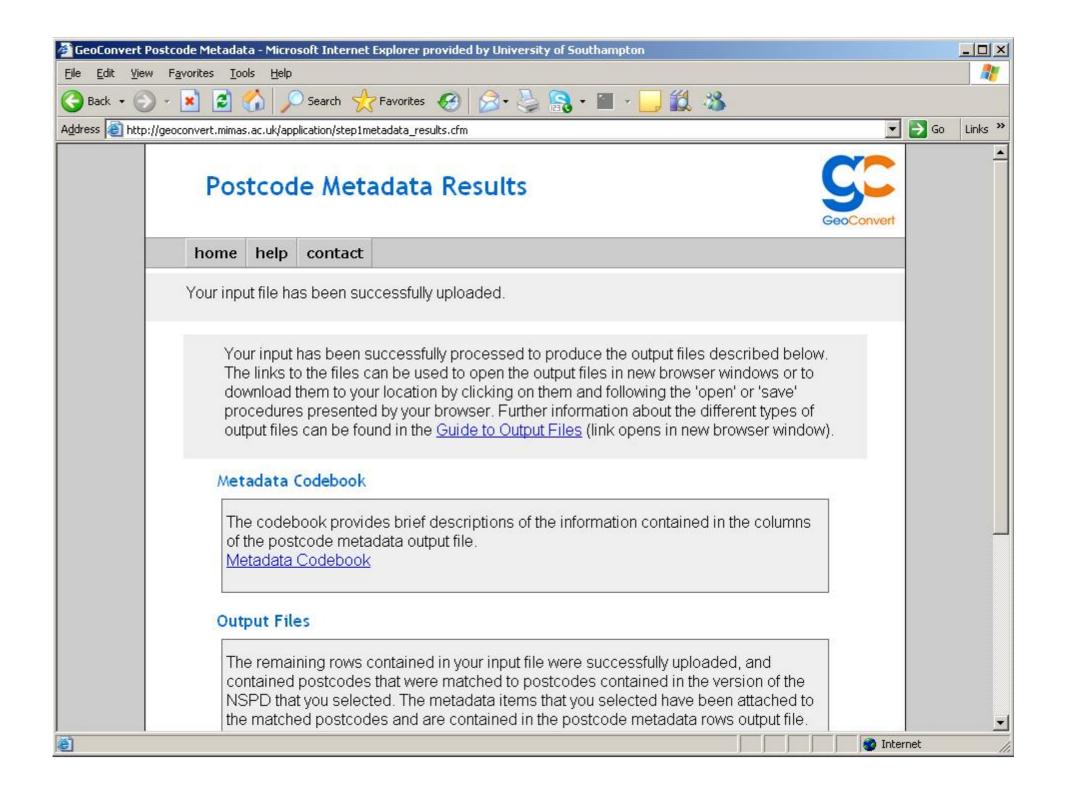


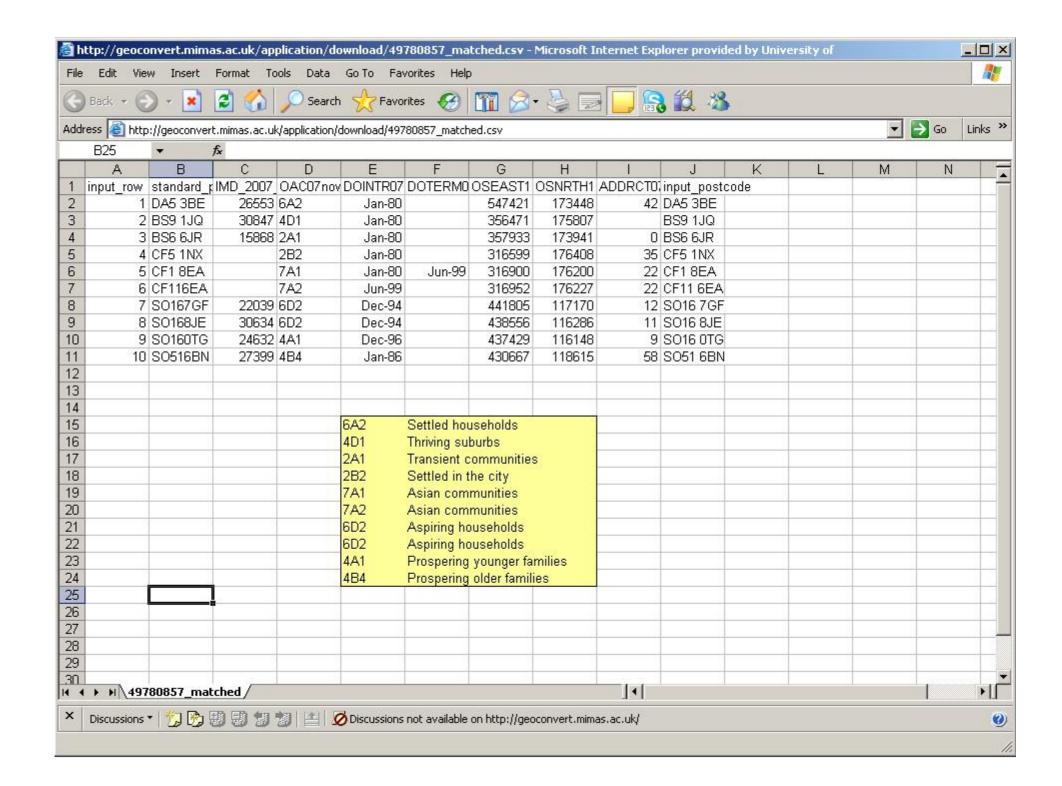


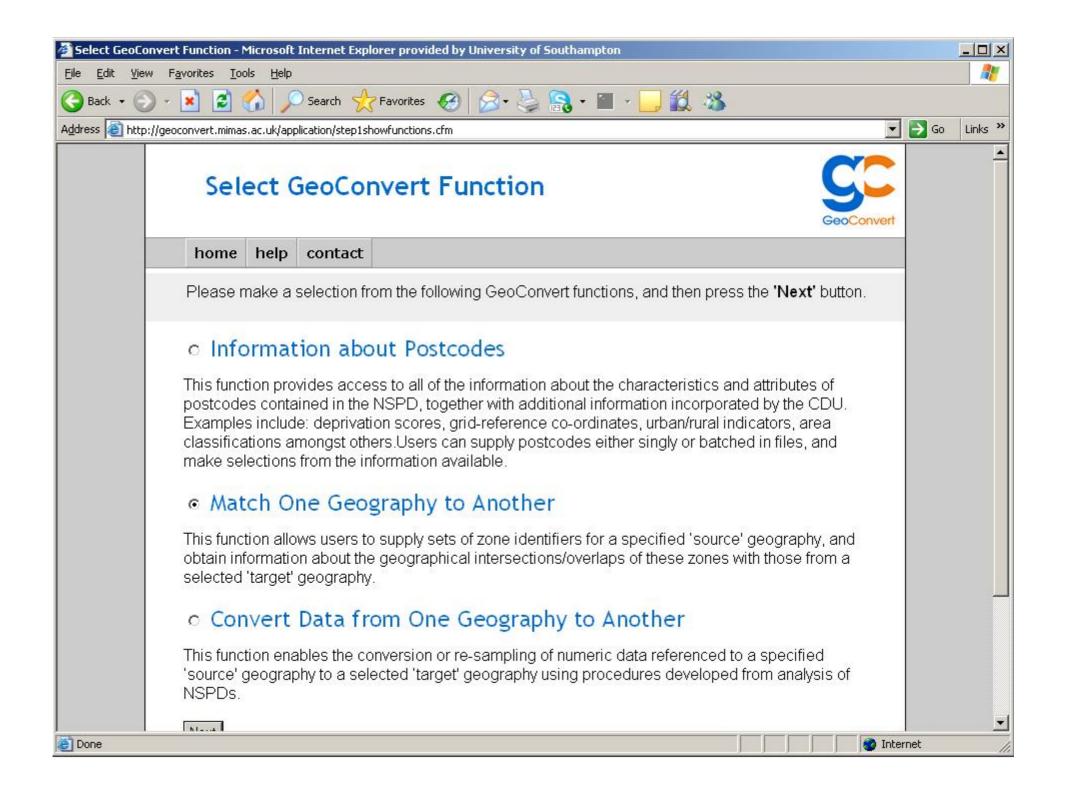


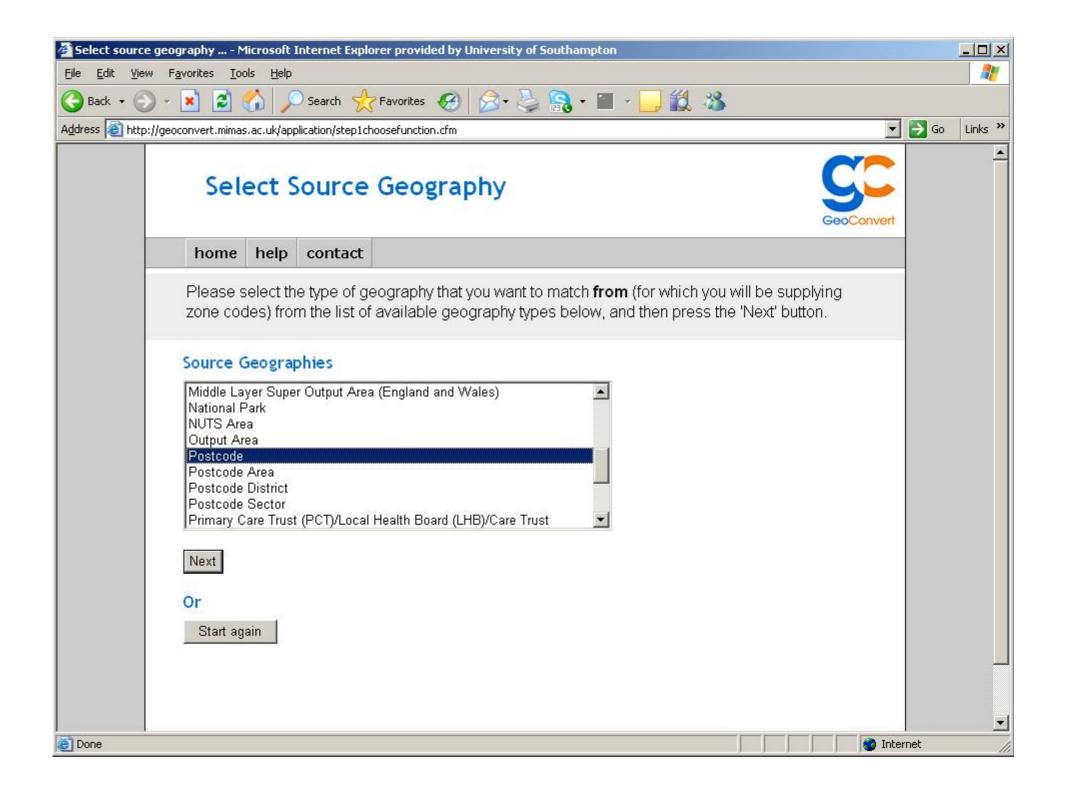


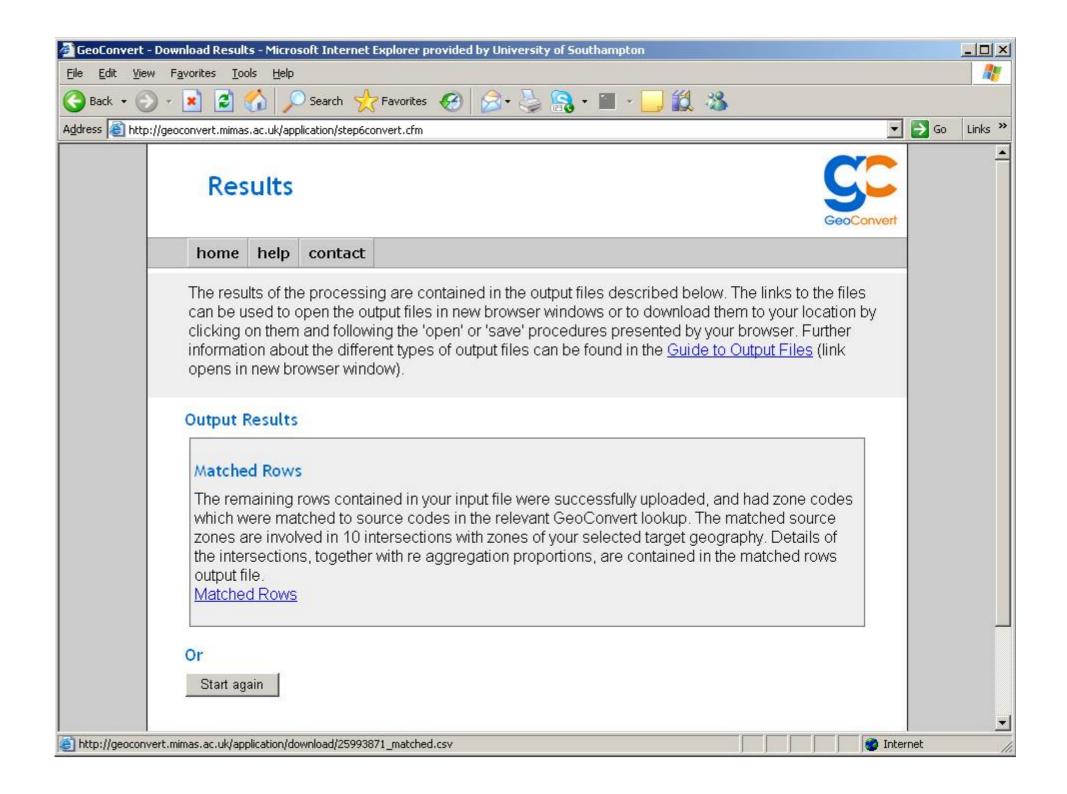


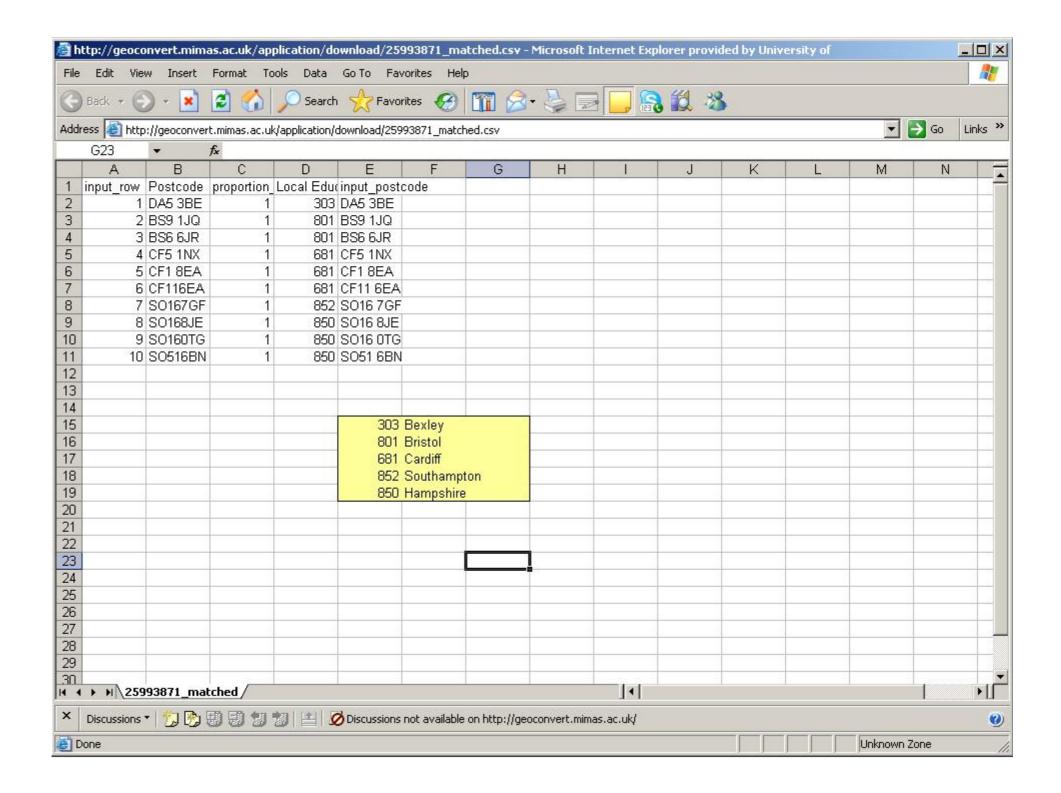


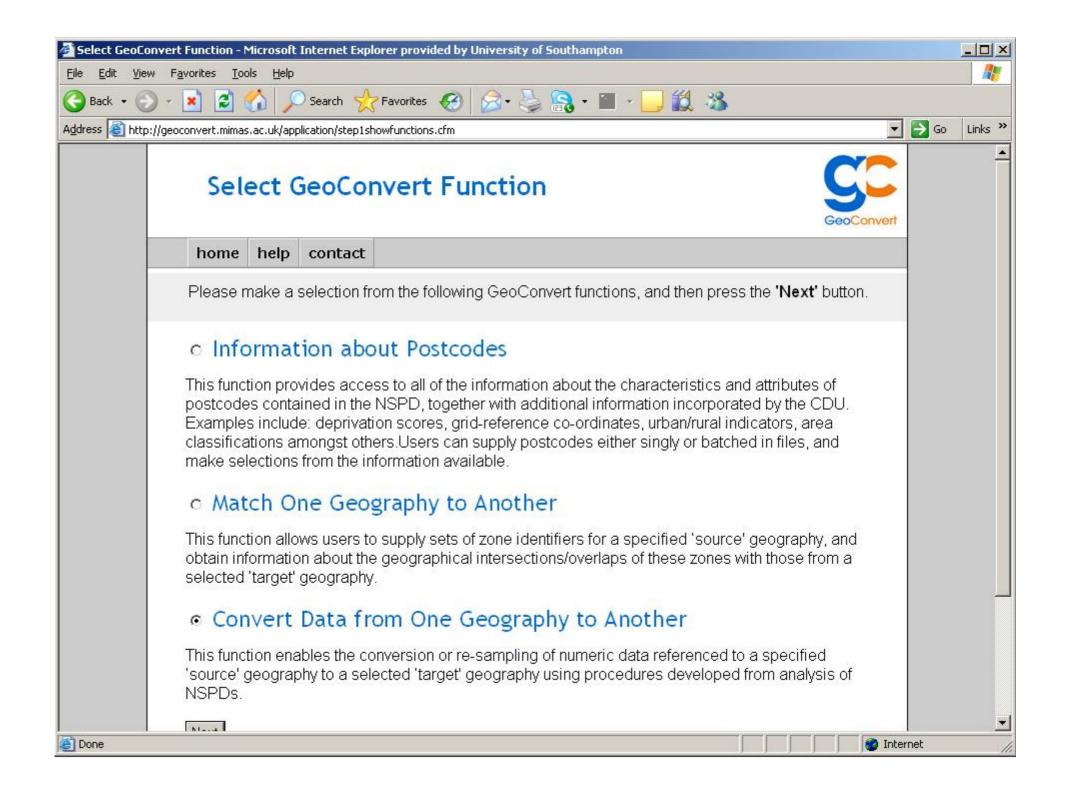


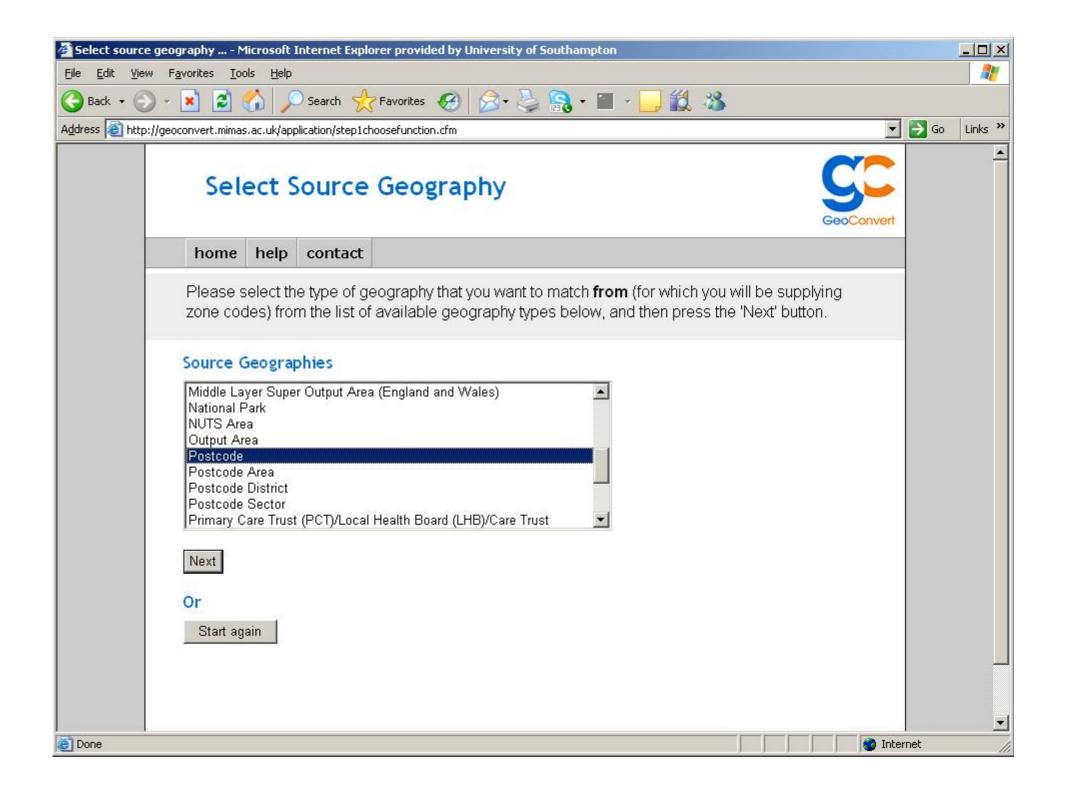


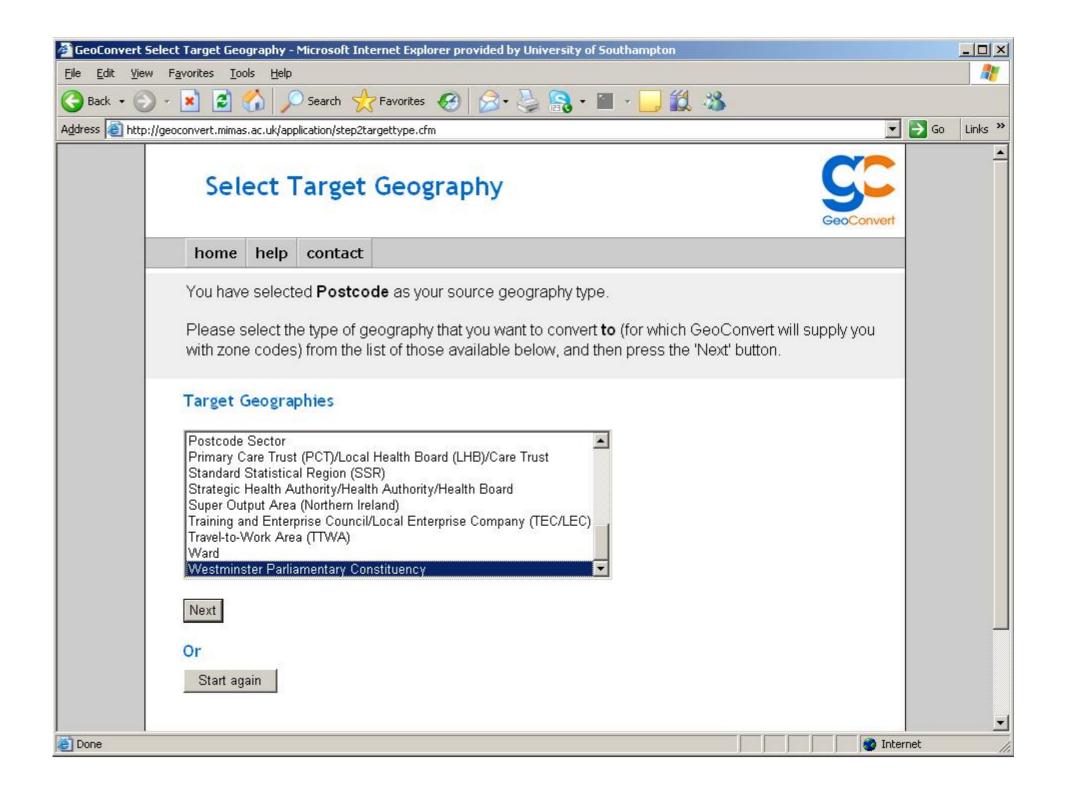


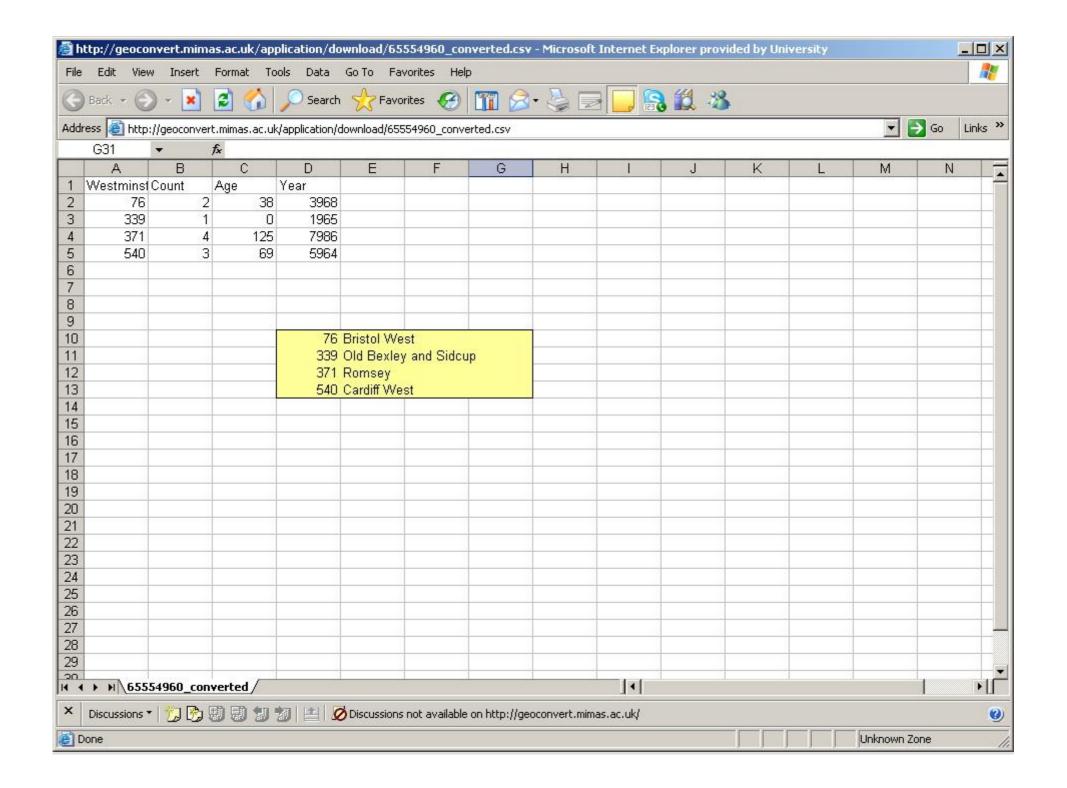










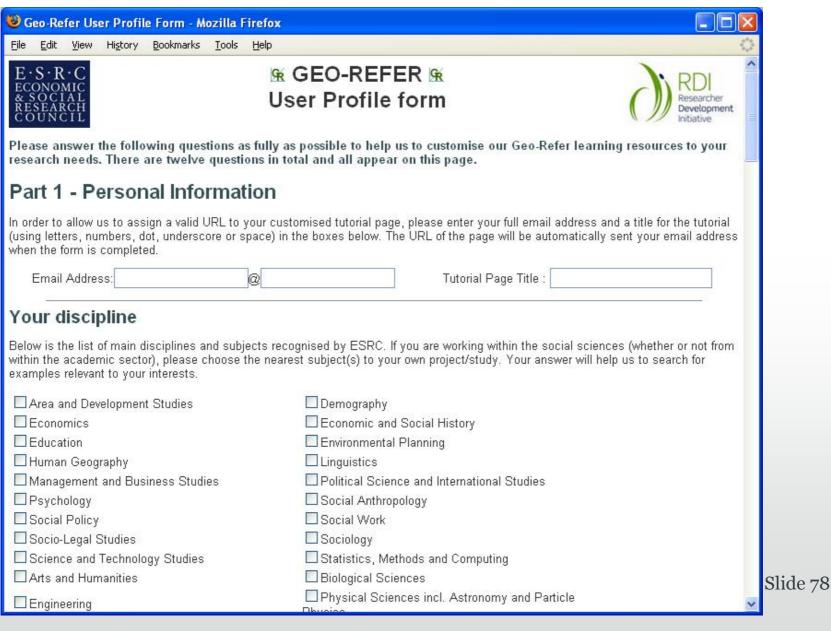


# Getting more help: how to use the Geo-Refer online resources

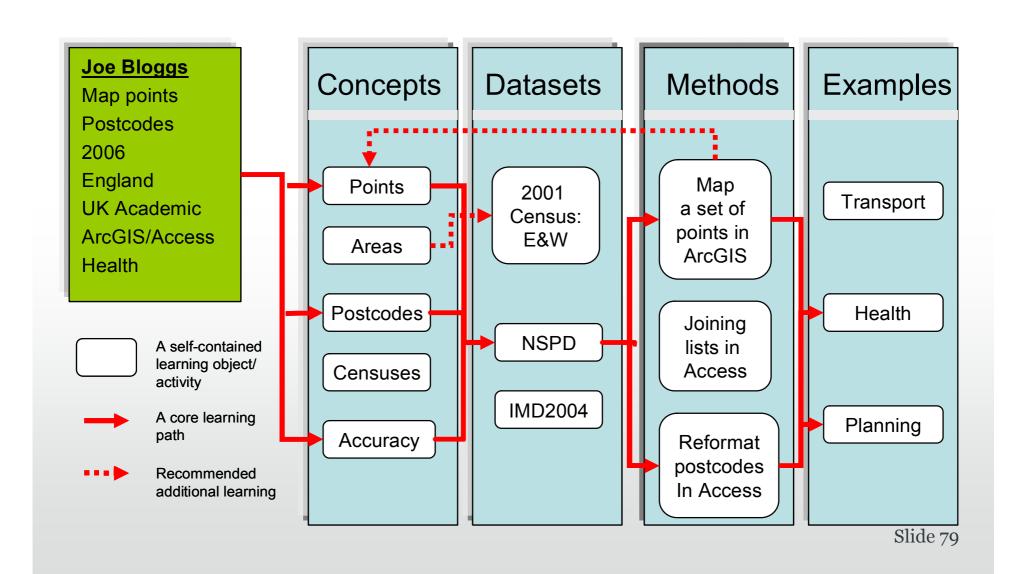
### Using the Geo-Refer resources

- URL <a href="http://www.geog.soton.ac.uk/geo-refer">http://www.geog.soton.ac.uk/geo-refer</a>
- Two options for using the site:
  - Complete the user profile form to specify own profile and Geo-Refer will best-match learning materials
  - Browse the resources, for an extensive list of concepts, methods, datasets and examples

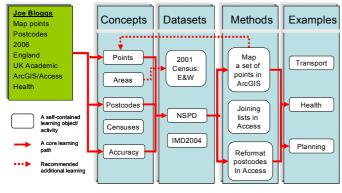
### Geo-Refer user profile form



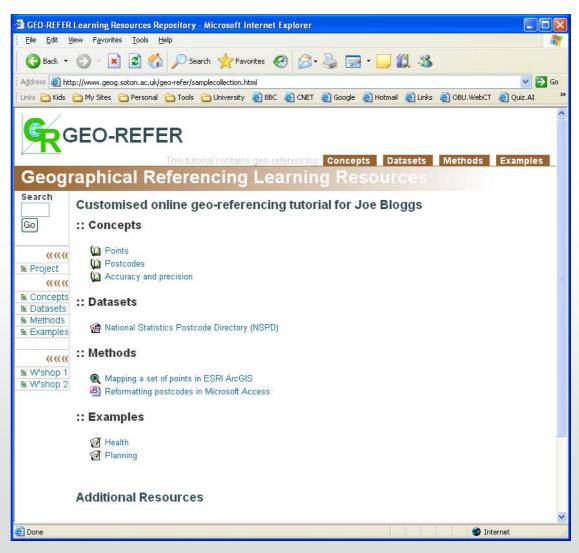
## Customised set of learning resources



### Customised online tutorial







# Really useful tools/datasets

# 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

### 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

#### Other useful sites...

Royal Mail Postcode Finder and Address Finder <a href="http://www.royalmail.co.uk">http://www.royalmail.co.uk</a>

> Registration required and limited number of uses

CASWEB <a href="http://census.ac.uk/casweb">http://census.ac.uk/casweb</a>

> 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 <u>www.earth.google.com</u> www.maps.live.com

> Street mapping and aerial photography (and many others)