Geographical referencing for social scientists: a practical workshop

21 November 2007

University of Edinburgh
Welcome and introductions

• (David Martin)
• Samantha Cockings
• Samuel Leung
• Workshop participants...
• Workshop URL:
  www.geog.soton.ac.uk/geo-refer/workshop3.html
Housekeeping

- Coffee and lunch
- Temporary login IDs
- Other facilities
- Travel assistance
Workshop structure

- A (very) little bit about the project
- Morning presentation of concepts and methods relevant to this group
- Online access to materials we have created so far
- Interactive afternoon workshop providing practice and advice on specific issues
Rules of engagement

• This is an informal workshop!

• Please interrupt…

• There is no such thing as a silly question
The Geo-Refer project

• ESRC Researcher Development Initiative
• Geographical referencing: a key data linkage tool
• Frequently needed by non-geographers but no obvious source of methodological guidance ("phone a friend?")
• Recent experience in creation of online learning materials
Georeferencing examples

- Link survey results to census data
- Associate postcoded patient records to indices of deprivation
- Allocate questionnaire responses to urban/rural neighbourhood types
- Identify which service delivery locations fall within different policy areas
- Use GPS to record survey locations
Researcher development...

• User needs (profile)
• Concepts (e.g. point and area objects)
• Datasets (e.g. postcode directory)
• Methods (e.g. tabular linkage)
• Examples (relevant to own needs!)
An example ...

I’ve got data from a GP survey. I’d like to map where the patients live.
Customised set of learning resources

User profile

Joe Bloggs
Map points
Postcodes
2006
England
UK Academic
ArcGIS/Access
Health

Concepts
- Points
- Areas
- Postcodes
- Censuses
- Accuracy

Datasets
- 2001 Census: E&W
- NSPD
- IMD2004

Methods
- Mapping a set of points in ArcGIS
- Joining lists in Access
- Reformatting postcodes in Access

Examples
- Transport
- Health
- Planning
Customised online tutorial
User profiling

• Face-to-face experience, but…
• Importance of understanding how users are encountering georeferencing problems
• Desire to structure materials appropriately
• Need to harvest ‘real world’ user examples that will be of help to others
• Help!
Please complete user profile form now!

[Image of Geo-Refer User Profile Form]

Part 1 - Personal Information

Please answer the following questions as fully as possible to help us to customise our learning resources to your research needs. There are eleven questions in total and all appear on this page.

In order to allow us to monitor usage of these resources, please enter your full email address in the box below.

Your discipline

Below is the list of subject areas used by ESRC. If you are working within the social sciences (whether or not form within the academic sector), please choose the nearest subject(s) to your own project/study. Your answer will help us to search for examples relevant to your interests.

- [ ] Area and development studies
- [ ] Demography
- [ ] Economic and social history
Workshop programme

• 11.15 Georeferencing concepts and methods
• 12.40 Using Geo-Refer online resources
• 13.00 Lunch
• 14.00 Hands-on workshop
• 15.30 Presentation of worked examples
• 16.00 Evaluation, tea and close
Georeferencing concepts
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: 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
Focusing on points and areas...

• Here focusing on point and area geographical references
• Centroid points frequently used to locate areas where boundaries are unavailable or uncertain
• Maybe geometric, population-weighted, other...
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…
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
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 Manchester
- **Precise location**
  - provides detail (but not necessarily accurate)
  - e.g. respondent lives at 255b Stockport Road
- **Need to assess and adopt different strategies according to purpose**
Map scale

- Ratio of a distance on the map to the corresponding distance on the ground

- Leads to inclusion/generalization of different features, boundary details, etc.
Map projections

- 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!
What about geographical information systems (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
- Lots of other functions not needed here…
- This is not a GIS course!
Georeferencing methods
Resources

• This session:
  – General overview
  – Worked examples

• This afternoon:
  – Website > Sample resources
  – Try them out!
Georeferencing methods

1 Data linkage
   – Tabular/Look-up tables
   – Spatial

2 Mapping
   – Points
   – Areas
Tabular data linkage example

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

Dataset 1

Asthma rates  GP surgeries

Dataset 2

Deprivation scores  Super Output Areas
Tabular data linkage example

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

<table>
<thead>
<tr>
<th>Asthma rates</th>
<th>GP surgeries</th>
<th>Postcodes of GP surgeries</th>
<th>Postcode to Super Output Area LUT</th>
<th>Deprivation scores</th>
<th>Super Output Areas</th>
</tr>
</thead>
</table>

Dataset 1

Georeferences

Geographical lookup table (LUT)

Dataset 2
Tabular linkage tools

• Access, SPSS etc. – general purpose database or statistical packages for matching lists with common data items

• NB. GeoConvert – online tool specific to recent UK postcode-based lookup
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!
Why??
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

1 Data linkage
   – Tabular/Look-up tables
   – Spatial

2 Mapping
   – Points
   – Areas
Spatial data linkage example

Research question: What is the relationship between car crime and policing policies?

Police beat areas: not defined according to any recognised administrative geography

Car crimes: occur at many locations which cannot be addressed or postcoded
Spatial data linkage example

Spatial linkage requires intersection of area and point coordinates to associate crimes with beat areas

Police beat areas

Car crimes
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
Spatial linkage tools

• Spatial linkage usually done using a GIS
  – Various GIS software packages
  – ESRI (ArcGIS, ArcInfo, ArcView)
  – MapInfo
  – IDRISI

• Remember: This is not a GIS course!
• Current worked examples use ArcGIS
Allocation: joining poly data to points data

Car crimes: points

Police beats: areas

Spatial link?

Point-in-polygon

Allocate polygon attributes to points
Allocation: joining poly data to points data

Spatial join: polygons to points

Sample output
Aggregation: joining points data to poly data

Car crimes: points

Spatial link?

Point-in-polygon

Police beats: areas

Aggregate points and/or their attribute data to polygons
Aggregation: joining points data to poly data

Spatial join: points to polygons

Sample output
Georeferencing methods

1. Data linkage
   - Tabular/Look-up tables
   - Spatial

2. Mapping
   - Points
   - Areas
Methods Part 2: Mapping

• Map locations of points and/or areas
  – E.g. grid-refs of survey responses
  – E.g. boundaries of Local Authority Districts (LADs)

• Map attributes of areas
  – E.g. Link survey responses on attitudes to recycling to LADs then map response rates by LADs
Mapping a set of 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

How do I map these?
Mapping a set of grid-referenced points

Get data into format compatible with import to ArcGIS

- Grid-refs (.csv, .tab, .xls)
- Import into Access and define fields e.g. ID, X, Y
- Save in dBASE (IV) format (.dbf)

Add into ArcMap as Table

- Create event theme (Add XY data) using X and Y fields
- Export as shapefile

Import into ArcGIS

Get data into format compatible with import to ArcGIS

- Grid-refs (.csv, .tab, .xls)
- Import into Access and define fields e.g. ID, X, Y
- Save in dBASE (IV) format (.dbf)

Add into ArcMap as Table

- Create event theme (Add XY data) using X and Y fields
- Export as shapefile

Import into ArcGIS
Linking attribute data to boundary data

I’ve got boundary data and attribute data but how do I link them together?
Linking attribute data to boundary data

Define common field in attribute table and in shapefile attribute table

JOIN attributes from a table (based on BEATID)

Attributes linked to boundaries
Creating a choropleth/thematic map

- OK, I’ve got my attributes linked to boundaries, but how do I map them?

- Choropleth/thematic/area-shaded maps
- Categorical or continuous data?
- Be careful about choice of classification scheme, class breaks and colour schemes
Creating a choropleth/thematic map

- Add layer (shapefile)
- Select variable to map
- Select data type
- (Select classification scheme/class breaks)
- Select colour scheme
Creating a choropleth/thematic map: Categorical data

Map states of USA shaded by sub-region
Creating a choropleth/thematic map: Continuous data

Map states of USA shaded by population density
Really useful datasets/tools
Some really useful datasets/resources…

• National Statistics Postcode Directory
• UKBORDERS Digital Boundary Data
• Neighbourhood Statistics Services

• GeoConvert (based on NSPD)
• ONS Beginners Guide to UK Geography
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

www.edina.ac.uk/ukborders
# UKBORDERS Digital Boundary Data

<table>
<thead>
<tr>
<th>Coordinates of area boundaries downloadable in various GIS and mapping formats</th>
<th>Need to understand which zones are needed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primarily census-derived, covering whole UK</td>
<td>Need suitable software to load the boundaries</td>
</tr>
<tr>
<td>Census, administrative, health and electoral geographies</td>
<td>Large data volumes and further manipulation often required</td>
</tr>
<tr>
<td>Includes some historical (pre-1971 boundary sets e.g. 1951 local government)</td>
<td>No attribute data</td>
</tr>
</tbody>
</table>

Downloadable from UKBORDERS; registration required

[www.edina.ac.uk/ukborders](http://www.edina.ac.uk/ukborders)
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

www.neighbourhood.statistics.gov.uk
<table>
<thead>
<tr>
<th>Other useful sites…</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Royal Mail Postcode Finder and Address Finder</td>
<td>Registration required and limited number of uses</td>
</tr>
<tr>
<td>Digimap – Ordnance Survey mapping, inc. historical</td>
<td>JISC-funded subscription service at Edina</td>
</tr>
<tr>
<td>Google Earth and Microsoft Live Search Maps</td>
<td>Street mapping and aerial photography (and many others)</td>
</tr>
</tbody>
</table>

- [www.royalmail.co.uk](http://www.royalmail.co.uk)
- [www.edina.ac.uk/digimap](http://www.edina.ac.uk/digimap)
- [www.maps.live.com](http://www.maps.live.com)
- [www.earth.google.com](http://www.earth.google.com)
Data linkage with GeoConvert

Developed by MIMAS

Online tools for NSPD extraction and matching

http://geoconvert.mimas.ac.uk/
GeoConvert functions

• Obtain information about postcodes
• Match one geography to another
• Convert data from one geography to another

• Type in codes or batch upload
• NB. 2006 postcodes + onwards only!
Using GeoConvert: general structure

- Specify version of directory to use
- Specify to and from geographies
- Specify other fields of interest
- Enter code or upload file
- Download matched file
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

www.statistics.gov.uk/geography/beginners_guide.asp
Using Geo-Refer online resources
Plan learning activities for this afternoon

• Review:
  – Your customised tutorial
  – Full index of available Geo-Refer on-line resources
  – What you’ve heard/seen this morning

• Identify what you’d like to work on (individually or in group(s)) and let us know
Lunchtime!
Practical workshop session
Datasets provided...

• Quality and outcomes framework 2004/5 data for general practitioners
  – Achievement data at general practice level
  – Practice code-name lookup table
    www.ic.nhs.uk/services/qof/data/

• Link practices to postcodes, postcodes to locations, locations to map...
Presentation of examples
Workshop evaluation form

• Immediate feedback on workshop
• We would like to find out more about your projects so that we can use them as examples to help others…
• Thank you!
Combining Census Aggregate Statistics with Digital Boundary Data for Mapping and Spatial Analysis

UK Censuses provide an unrivalled source of contemporary and historical demographic and socio-economic information, available with complete UK coverage at a range of geographical scales down to small local areas. Census information has a vast range of actual and potential uses, both on its own, and in conjunction with information from other sources.

The course begins with a brief overview of the wide range of information and resources available from the units of the ESRC Census Programme, and then focuses on the census aggregate statistics, the most widely used component of the census outputs, available from the Census Dissemination Unit (CDU), and the digital boundary data (DBD) used for computer mapping and spatial analyses, available from UKBORDERS. The course is aimed primarily at postgraduates, researchers, teaching staff and data librarians, and consists of a mixture of presentations and supervised practicals including exercises. It aims to be flexible enough to accommodate the particular interests of attendees, whilst providing answers to common questions such as:

- What are the census aggregate statistics?
- What is digital boundary data?
- What are they used for, and how might I use them?
- How can I identify and obtain information of interest to me?
- How can I combine aggregate statistics and DBD to produce maps and perform spatial analyses?
- What methods could I use to combine non-census information (e.g. postcoded sample data) with aggregate statistics and DBD?

Places can be booked using the Miras online course booking form

http://www.mimas.ac.uk/courses/content.html#census