Geographical referencing for social scientists: a practical workshop
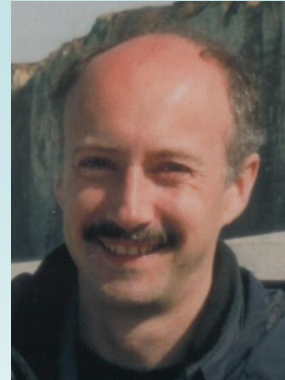
21 November 2007
University of Edinburgh

Welcome and introductions

- (David Martin)
- Samantha Cockings
- Samuel Leung
- Workshop participants…
- Workshop URL:

  www.geog.soton.ac.uk/geo-refer/workshop3.html

Geo-Refer

Housekeeping

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

Geo-Refer

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

Geo-Refer

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

Geo-Refer

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

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

RDI Researcher Development Initiative

Geo-Refer

Researcher development…

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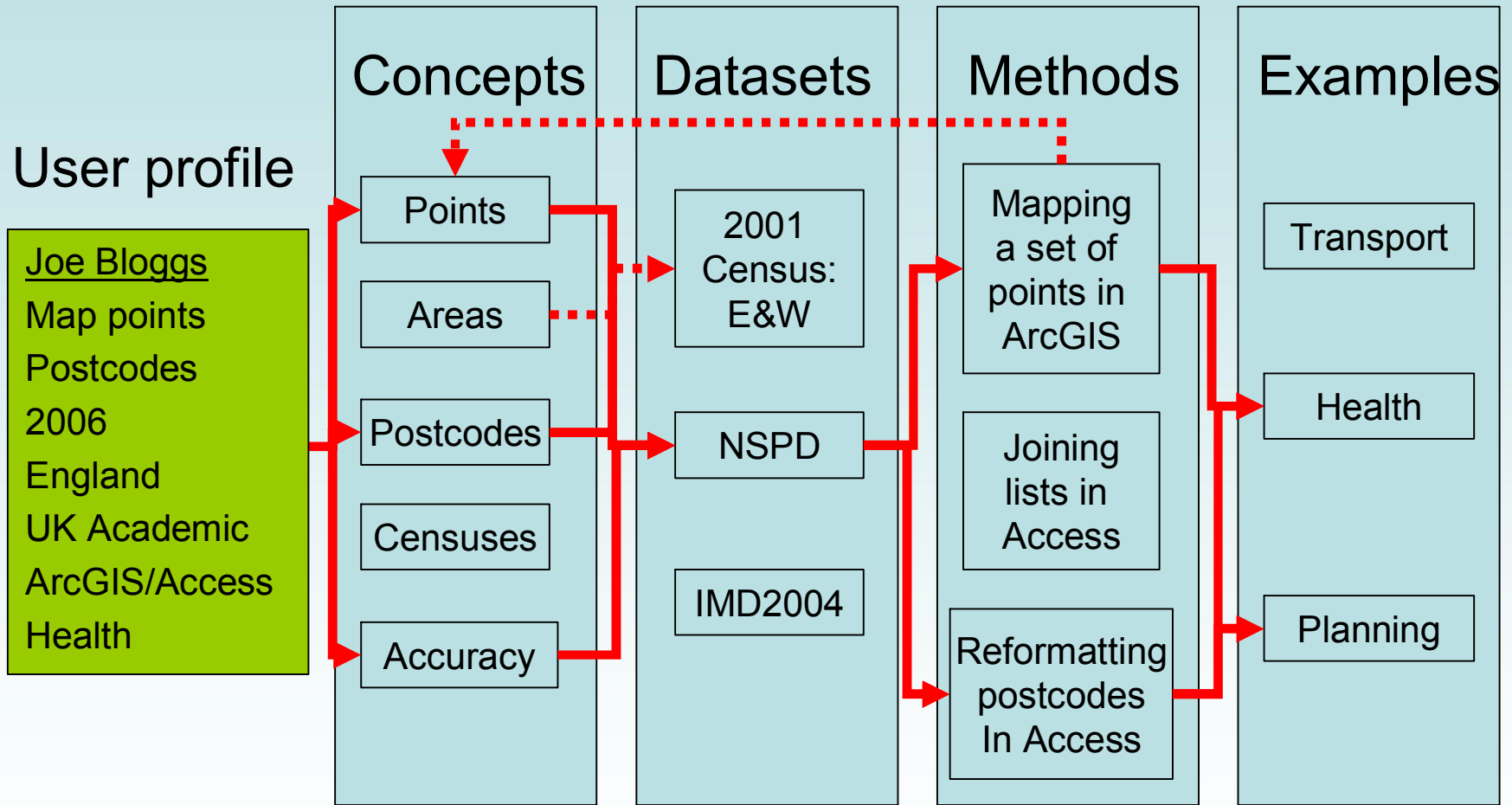- Examples (relevant to own needs!)
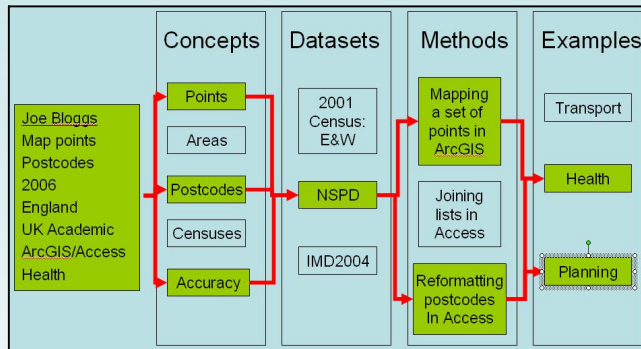
Geo-Refer

An example …

Joe Bloggs

I've got data from a GP survey. I'd like to map where the patients live.

Geo-Refer

Customised set of learning resources

Concepts

Datasets

Methods

Examples

User profile

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

Points

Areas

Postcodes

Censuses

Accuracy

2001 Census: E&W

NSPD

IMD2004

Mapping a set of points in ArcGIS

Joining lists in Access

Reformatting postcodes In Access

Transport

Health

Planning

Geo-Refer

Customised online tutorial
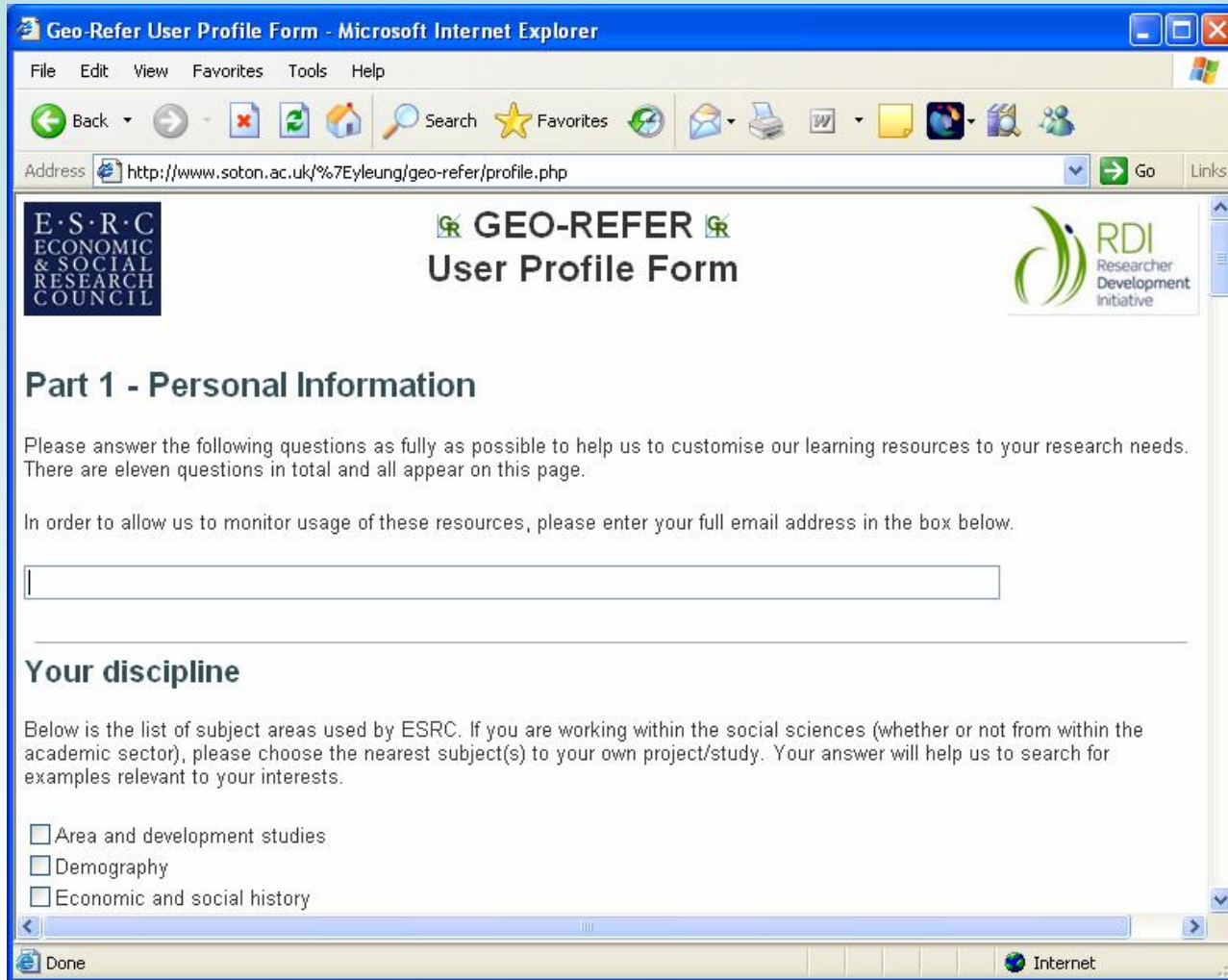
User profiling

- 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!

Geo-Refer

Please complete user profile form now!

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

Geo-Refer

Georeferencing concepts

ESRC ECONOMIC & SOCIAL RESEARCH COUNCIL

RDI Researcher Development Initiative

Geo-Refer

Geographical referencing of social science phenomena: 1

- Characteristics of people or events, recorded at mail addresses

- Geographical location of mail address

Photo: © David Martin

Geo-Refer

Geographical referencing of social science phenomena: 2

- Administration or policy related to defined areas, aggregations of individuals

- Geographical location of area boundaries

Photo: © David Martin

ESRC ECONOMIC & SOCIAL RESEARCH COUNCIL

RDI Researcher Development Initiative

Geo-Refer

Geographical referencing of social science phenomena: 3

- Phenomena that are linear, describing routes or flows

- e.g. bus route, telephone conversation, social network link

Photo: © David Martin

Photo: © David Martin

Geo-Refer

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

Photo: © David Martin

Geo-Refer

Geographical referencing of social science phenomena: 5

- Complex phenomena not captured by a simple location

- e.g. business locations, catchment populations

Photo © David Martin

Geo-Refer
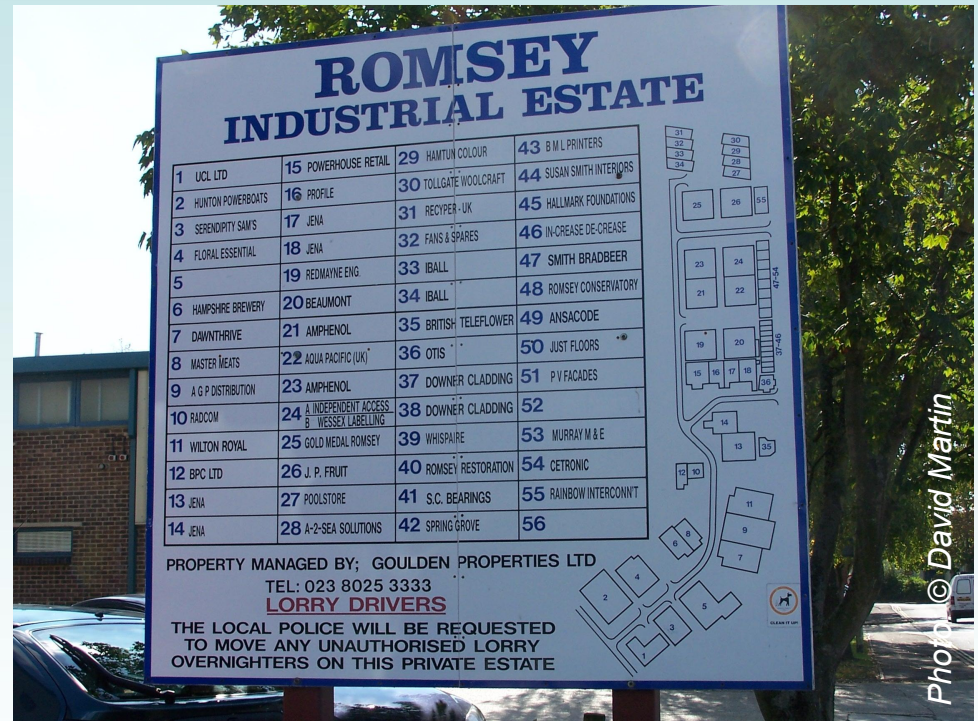
Geographical object types

- 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

Geo-Refer

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…

Geo-Refer

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

Geo-Refer

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**

Geo-Refer
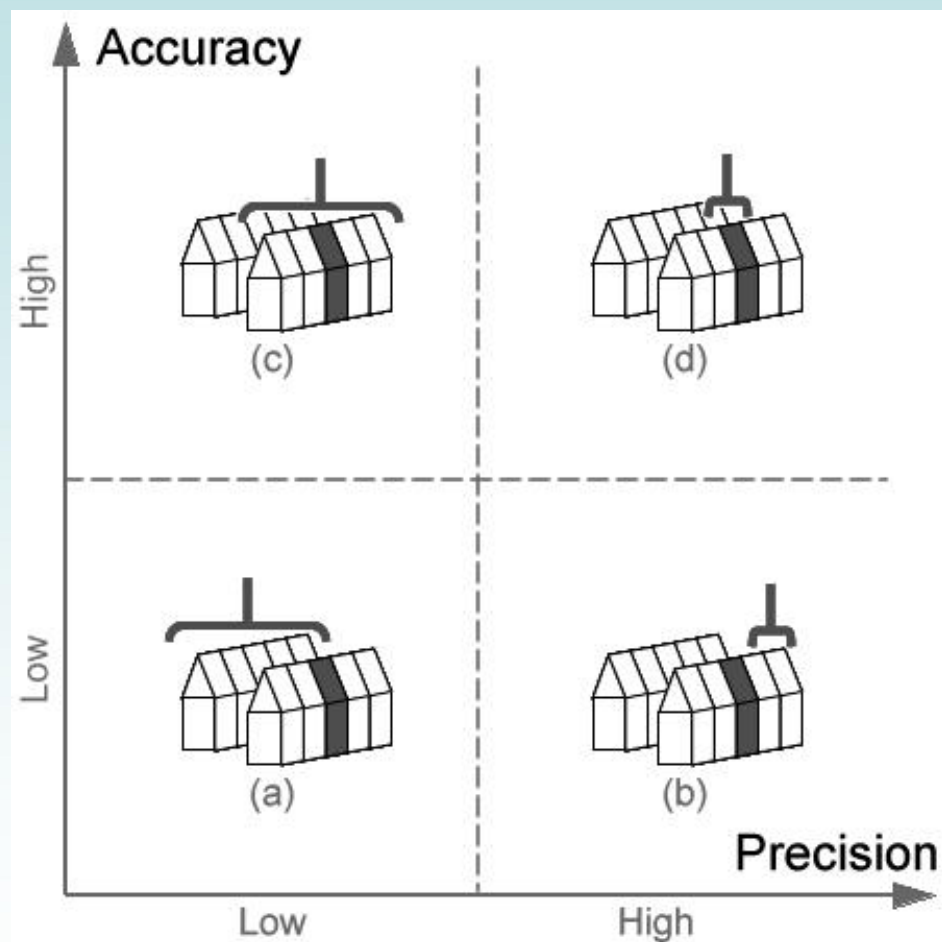
Map scale

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

(a) 1:25,000 (or 1/25,000)

(b) One centimetre to two hundred and fifty metres

(c) 0   5km

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

Map projections

(a) Map projection → (b) Main types

(i) Conformal

(ii) Equidistant

(iii) Equivalence
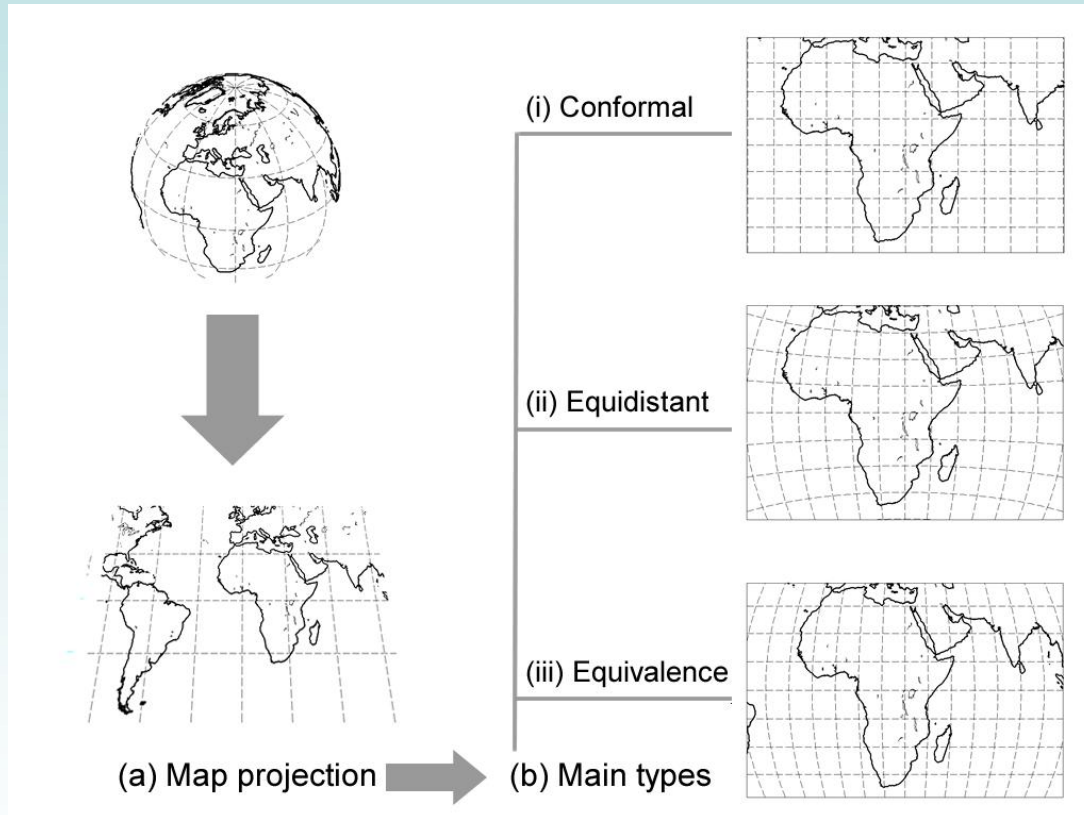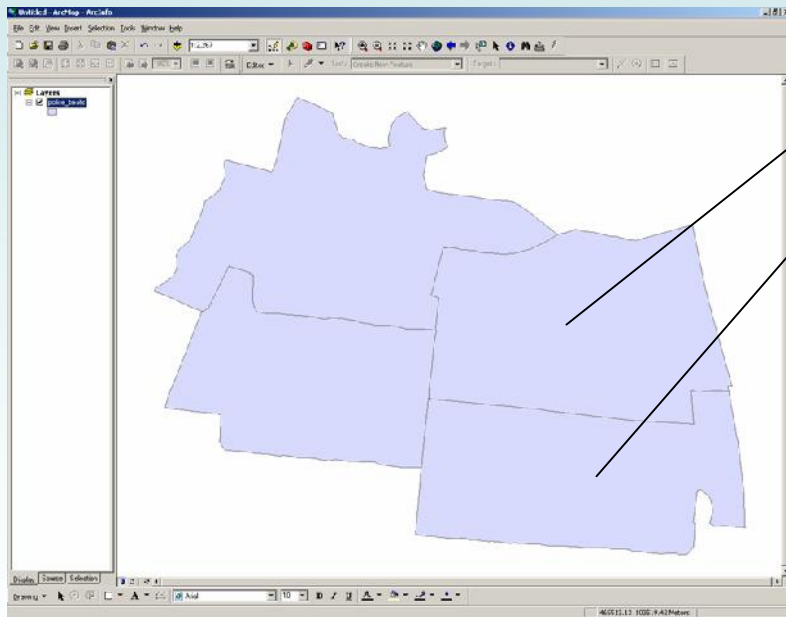
- Very important when using lat/long datasets

- 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!

Geo-Refer

Georeferencing methods

Resources

- **This session:**
  - General overview
  - Worked examples

- **This afternoon:**
  - Website > Sample resources
  - Try them out!

Geo-Refer
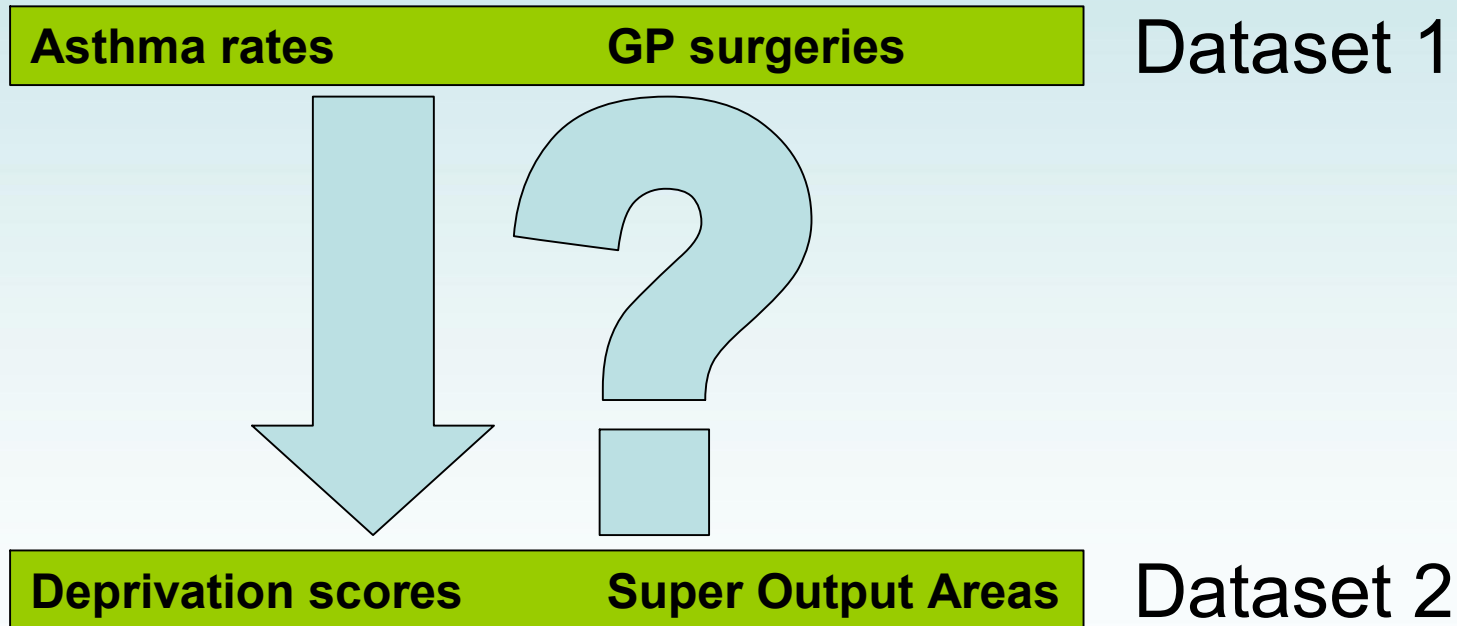
Georeferencing methods

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

2 Mapping
  – Points
  – Areas

Geo-Refer

Tabular data linkage example

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

| **Asthma rates** | **GP surgeries** | Dataset 1 |

| **Deprivation scores** | **Super Output Areas** | Dataset 2 |

Geo-Refer

Tabular data linkage example

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

| Asthma rates | GP surgeries |

Dataset 1

Postcodes of GP surgeries

Georeferences

Postcode to Super Output Area LUT

Geographical lookup table (LUT)

| Deprivation scores | Super Output Areas |

Dataset 2

Geo-Refer
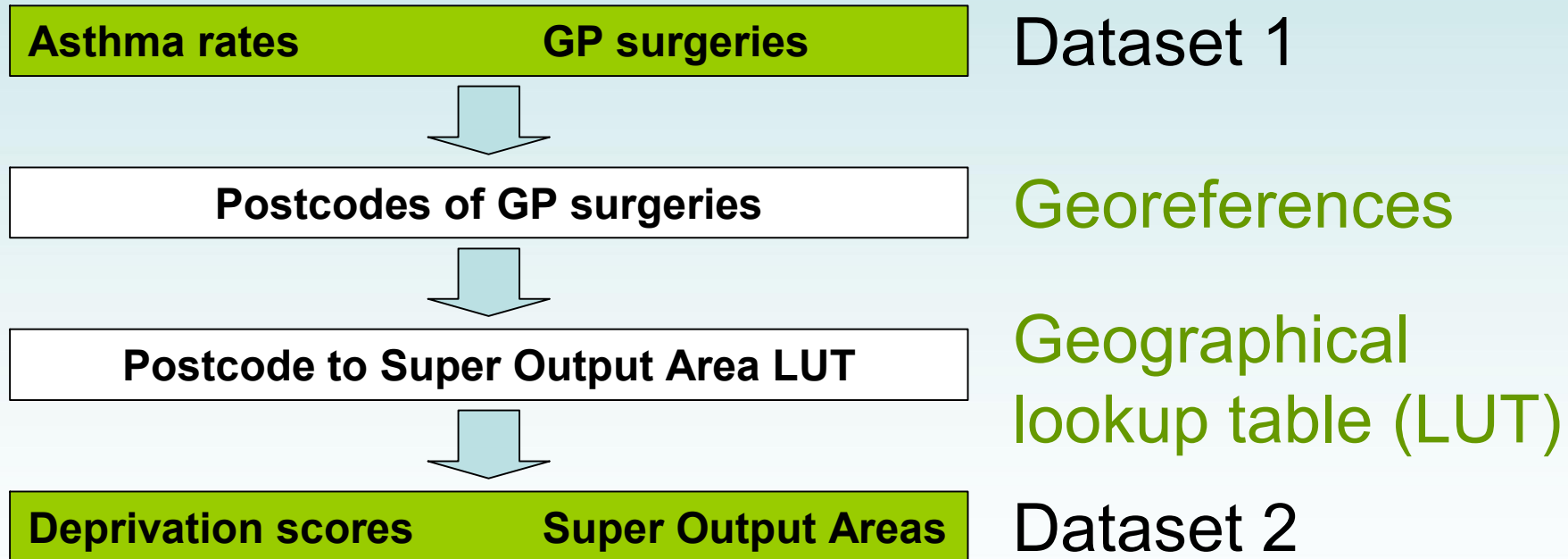
Tabular linkage tools

- Access, SPSS etc. – general purpose database or statistical packages for matching lists with common data items
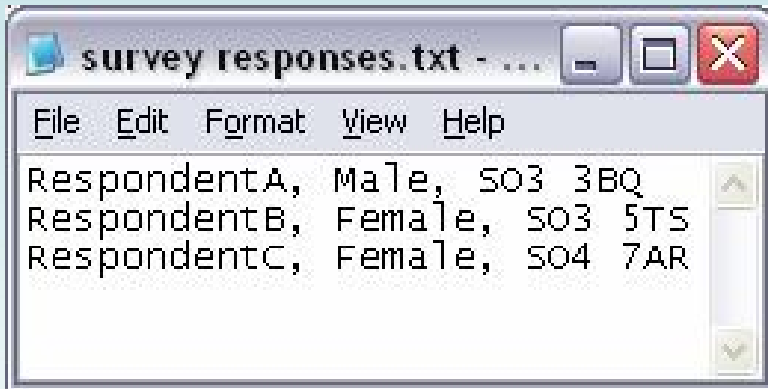
- NB. GeoConvert – online tool specific to recent UK postcode-based lookup

Data linkage example: input tables

Coded survey responses

```
survey responses.txt - ...
File  Edit  Format  View  Help
RespondentA, Male, SO3 3BQ
RespondentB, Female, SO3 5TS
RespondentC, Female, SO4 7AR
```
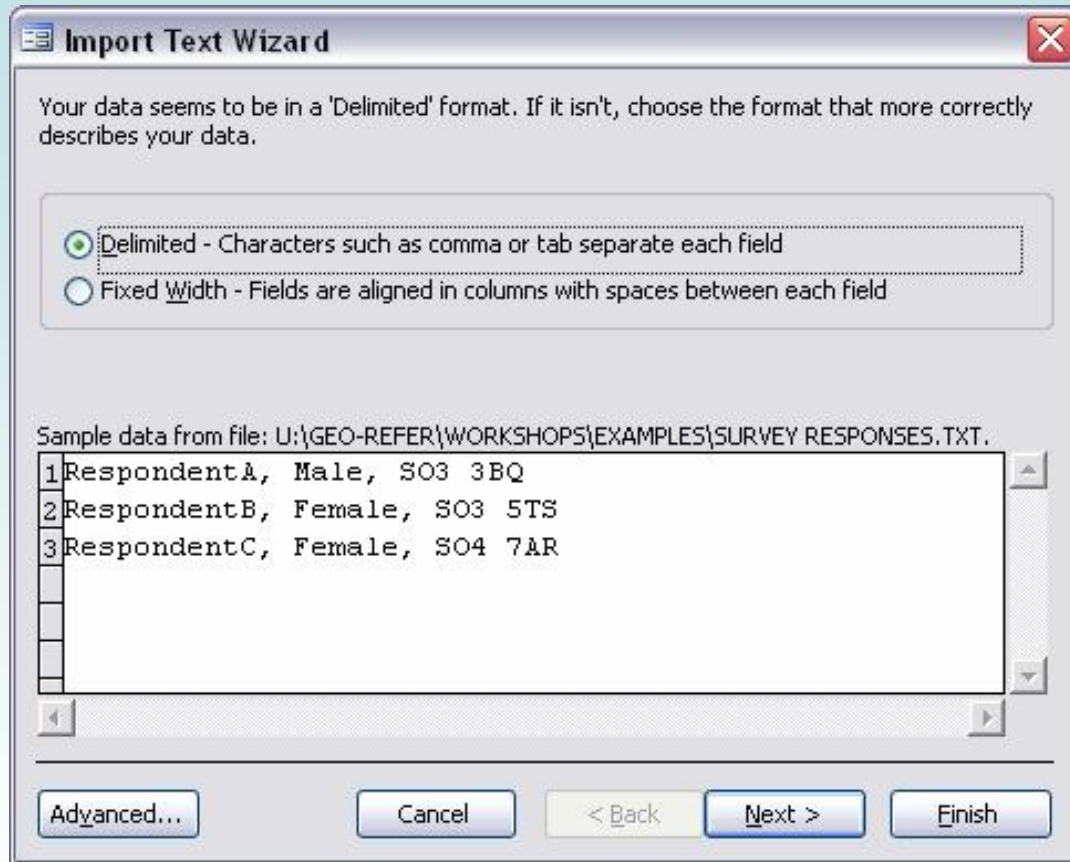
Respondent ID

Gender

Postcode

Postcode directory

```
postcode directory.txt ...
File  Edit  Format  View  Help
SO3  3BPWARD53
SO3  3BQWARD54
SO3  3BRWARD54
SO3  5TSWARD61
```

Postcode

Ward code

Import external data files

Import Text Wizard

Your data seems to be in a 'Delimited' format. If it isn't, choose the format that more correctly describes your data.

- ◉ Delimited - Characters such as comma or tab separate each field
- ○ Fixed Width - Fields are aligned in columns with spaces between each field

Sample data from file: U:\GEO-REFER\WORKSHOPS\EXAMPLES\SURVEY RESPONSES.TXT.

```
1 RespondentA, Male, SO3 3BQ
2 RespondentB, Female, SO3 5TS
3 RespondentC, Female, SO4 7AR
```

Advanced...   Cancel   < Back   Next >   Finish
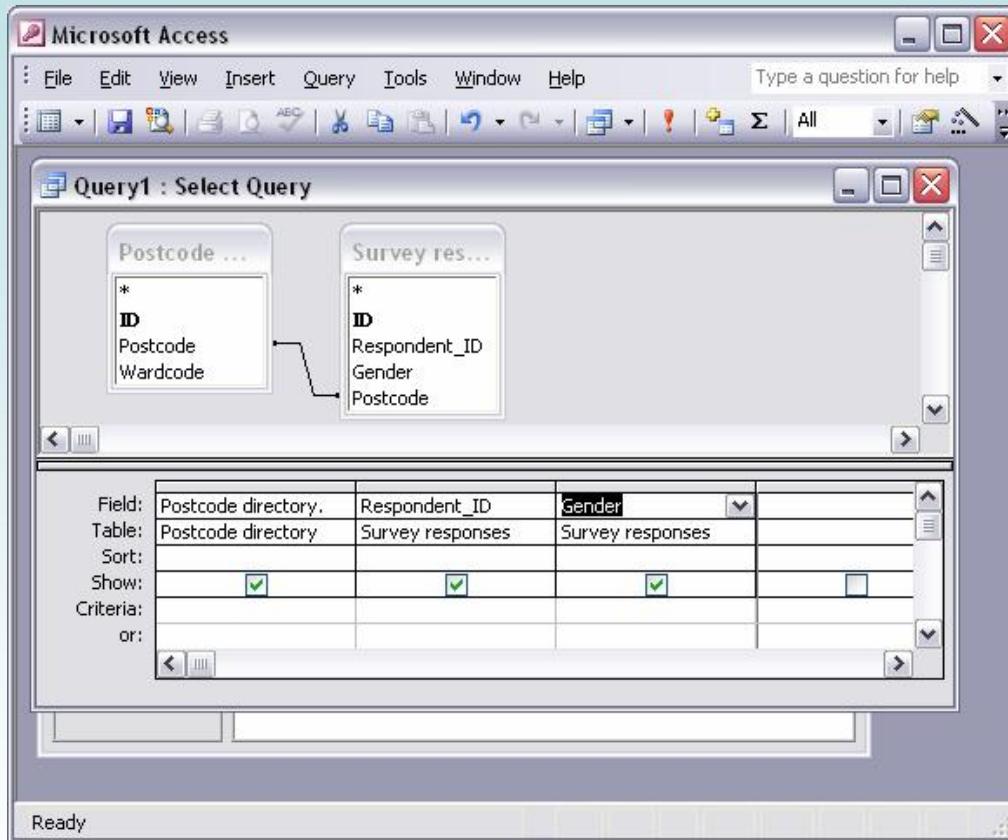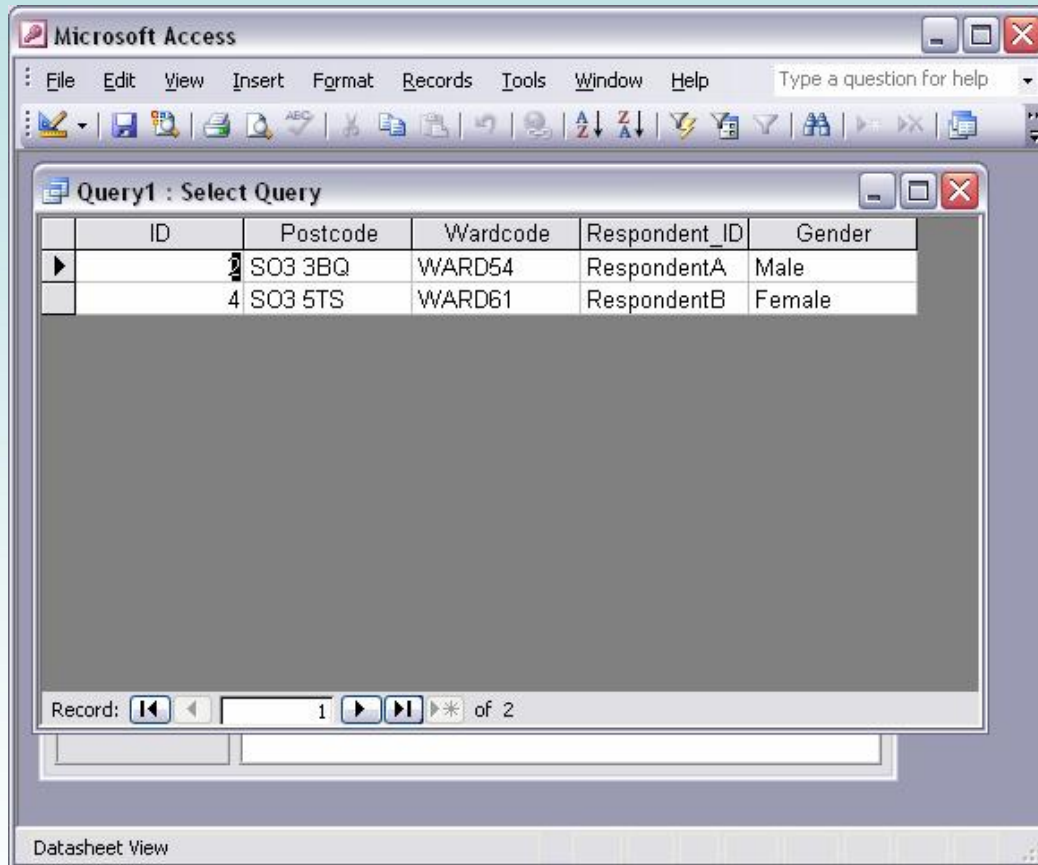
Get external data > import

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

Geo-Refer

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??

Geo-Refer

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!

Geo-Refer

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
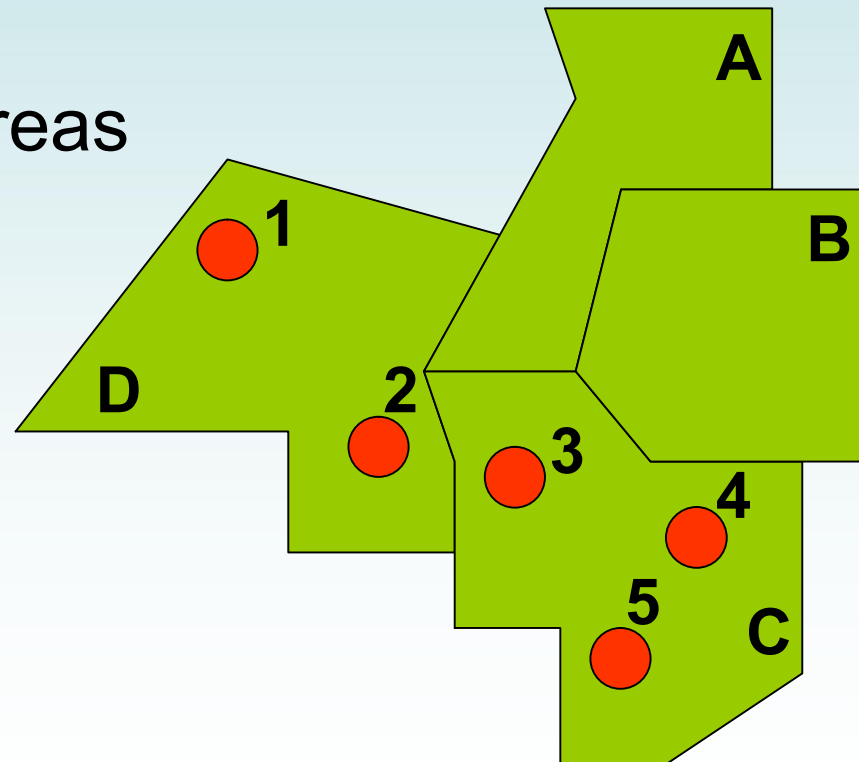
Geo-Refer

Spatial data linkage example

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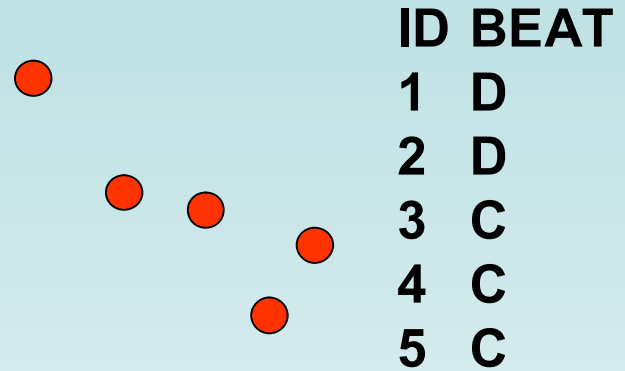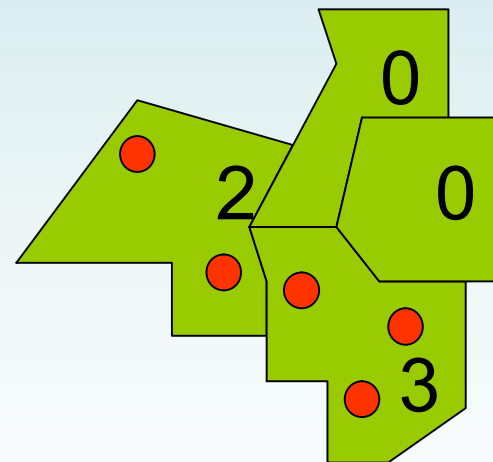
Police beat areas

Car crimes

A

B

D

1

2

3

4

5

C

RDI
Researcher
Development
Initiative

Geo-Refer

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

0

2   0

3

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

RDI
Researcher
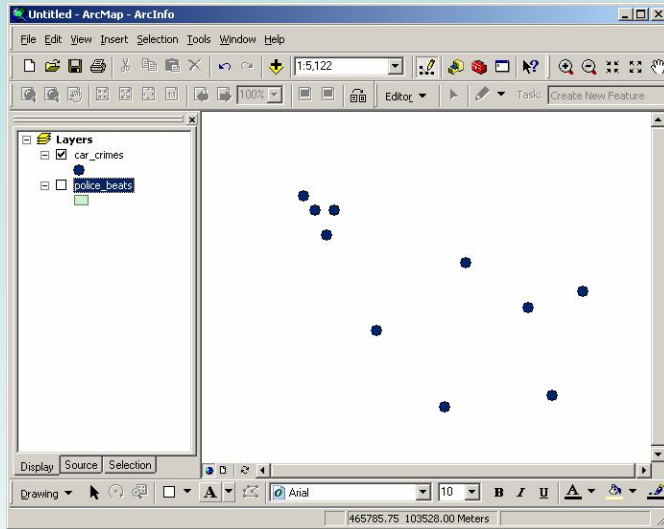Development
Initiative

Geo-Refer

Spatial linkage tools

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

- Remember: This is <u>not</u> a GIS course!
- Current worked examples use ArcGIS

Geo-Refer

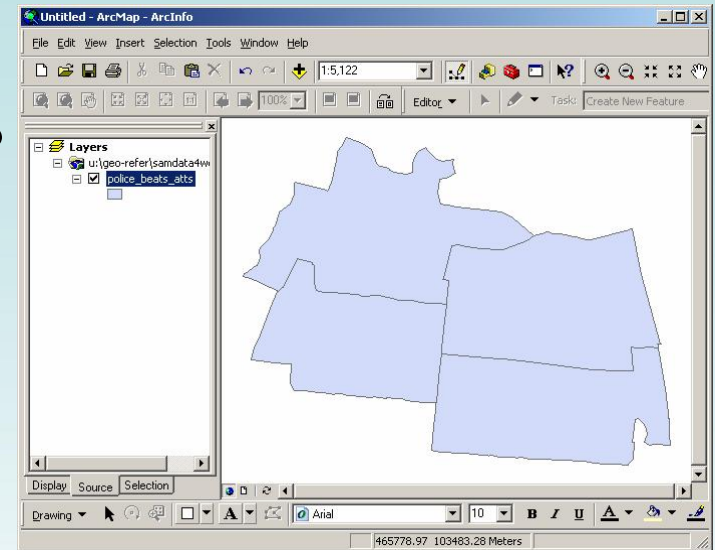# Allocation: joining poly data to points data

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
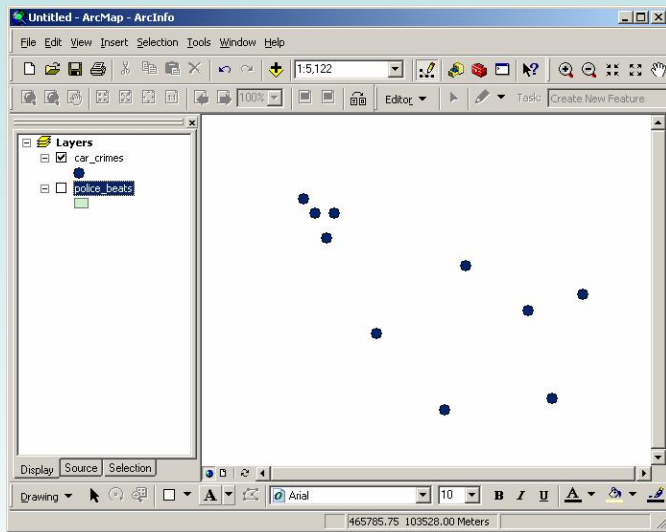
Police beats: areas

Spatial link?
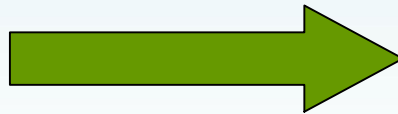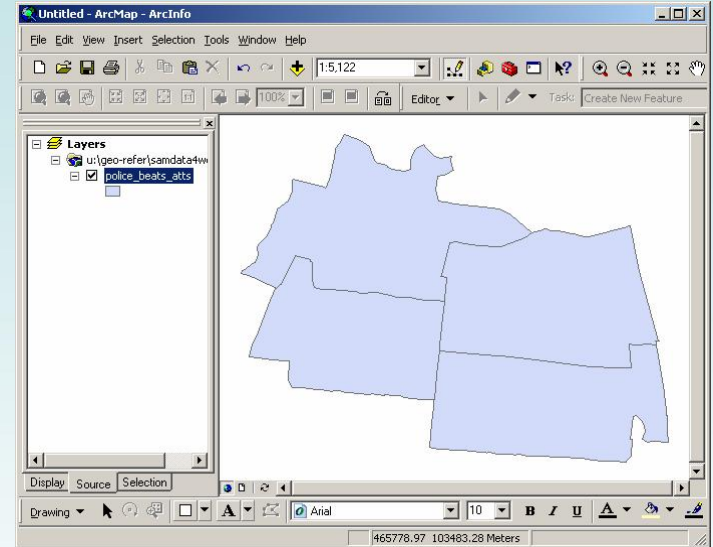
Point-in-polygon

Aggregate points and/or their attribute data to polygons

Geo-Refer

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

Geo-Refer

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?

Geo-Refer
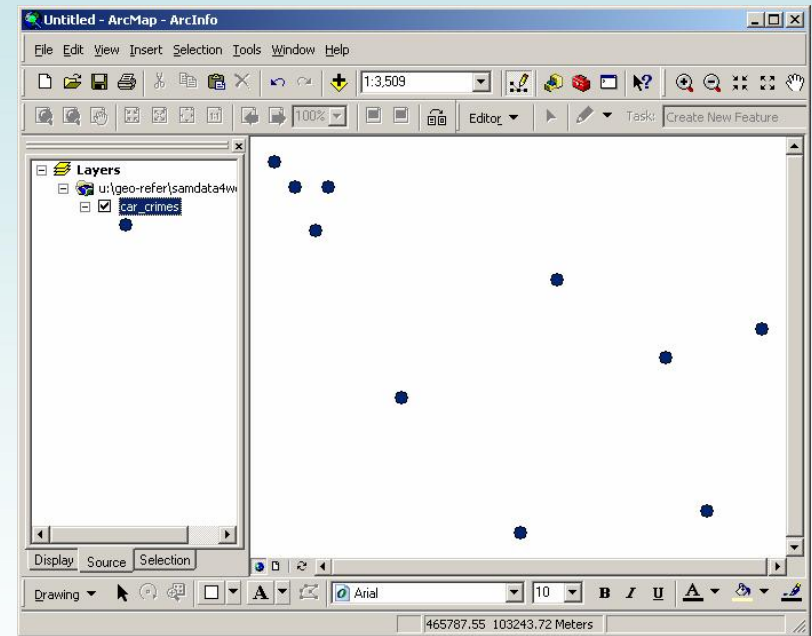
Mapping a set of grid-referenced points

Get data into format compatible with import to ArcGIS

Import into 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
```
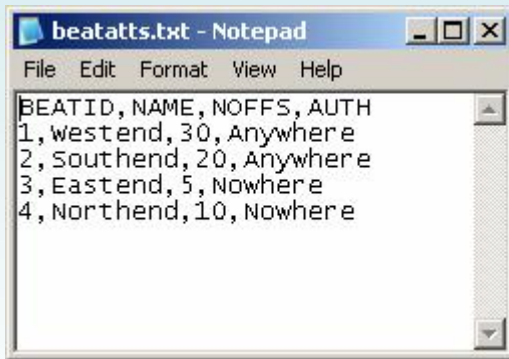

Add XY Data

A table containing X and Y coordinate data can be added to the map as a layer

Choose a table from the map or browse for another table:

CRIMES

Specify the fields for the X and Y coordinates:

X Field: X

Y Field: Y

Spatial Reference of Input Coordinates

Description:

Projected Coordinate System:
    Name: British_National_Grid

Geographic Coordinate System:
    Name: GCS_OSGB_1936

☐ Show Details

Edit...

OK   Cancel

Geo-Refer

Linking attribute data to boundary data

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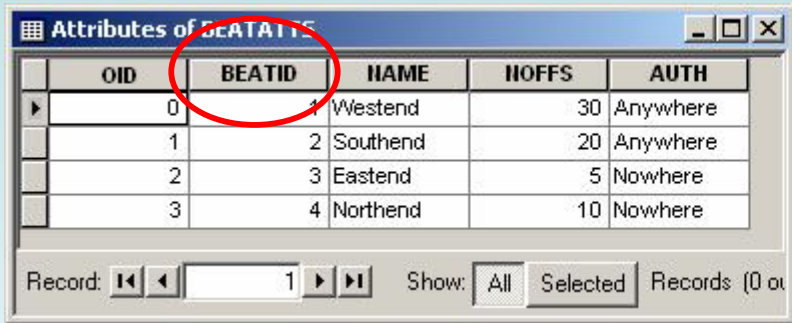
?
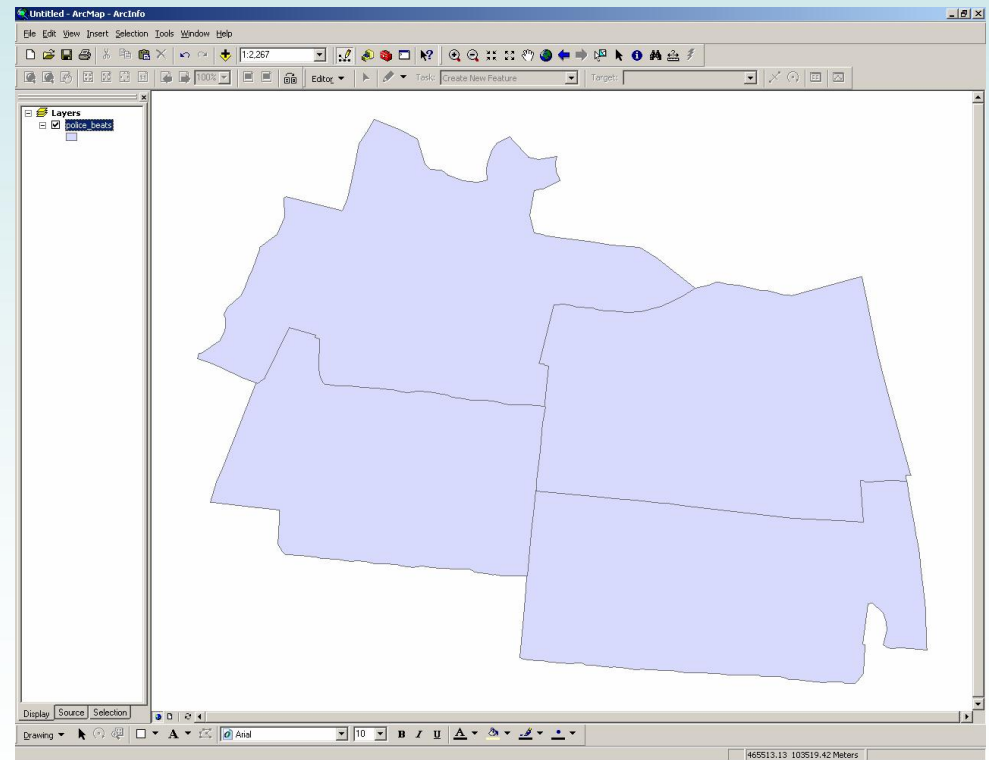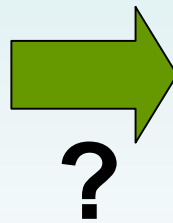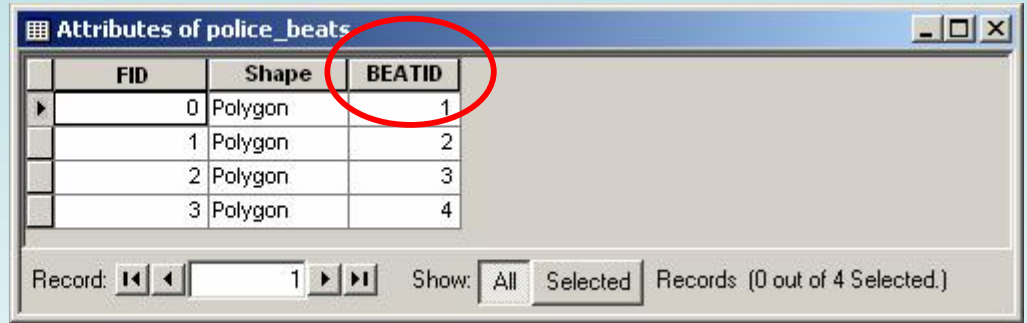
Linking attribute data to boundary data

Define common field in attribute table and in shapefile attribute table

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

Attributes of BEATATTS

Record: 1 Show: All Selected Records (0 ou...

FID	Shape	BEATID
0	Polygon	1
1	Polygon	2
2	Polygon	3
3	Polygon	4

Attributes of police_beats

Record: 1 Show: All Selected Records (0 out of 4 Selected.)

JOIN attributes from a table (based on BEATID)

Attributes of police_beats_atts

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

Record: 0 Show: All Selected Records (0 out of 4 Sel...

Attributes linked to boundaries

RDI Researcher Development Initiative

Geo-Refer

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

Geo-Refer

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

Geo-Refer
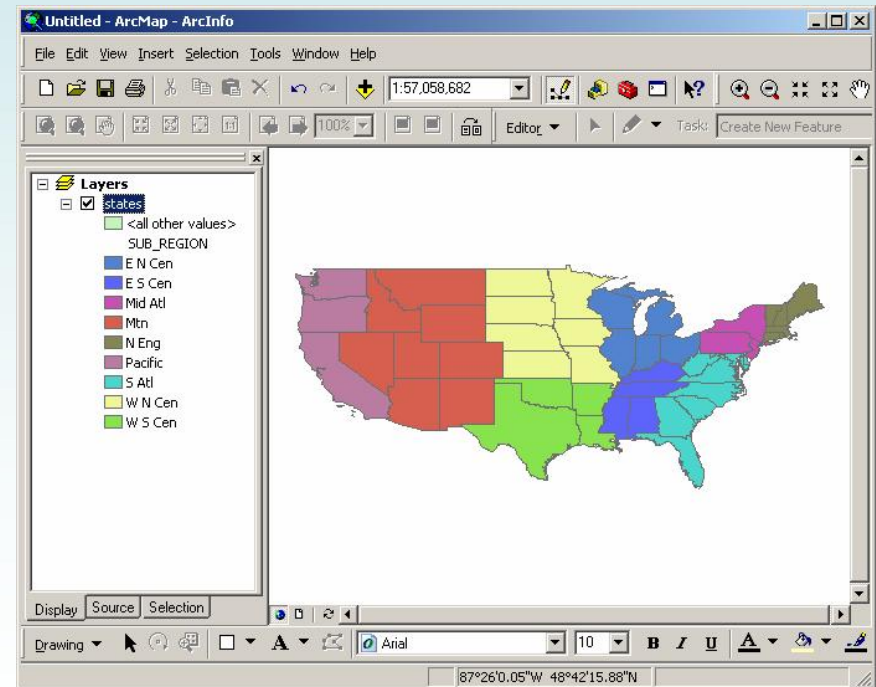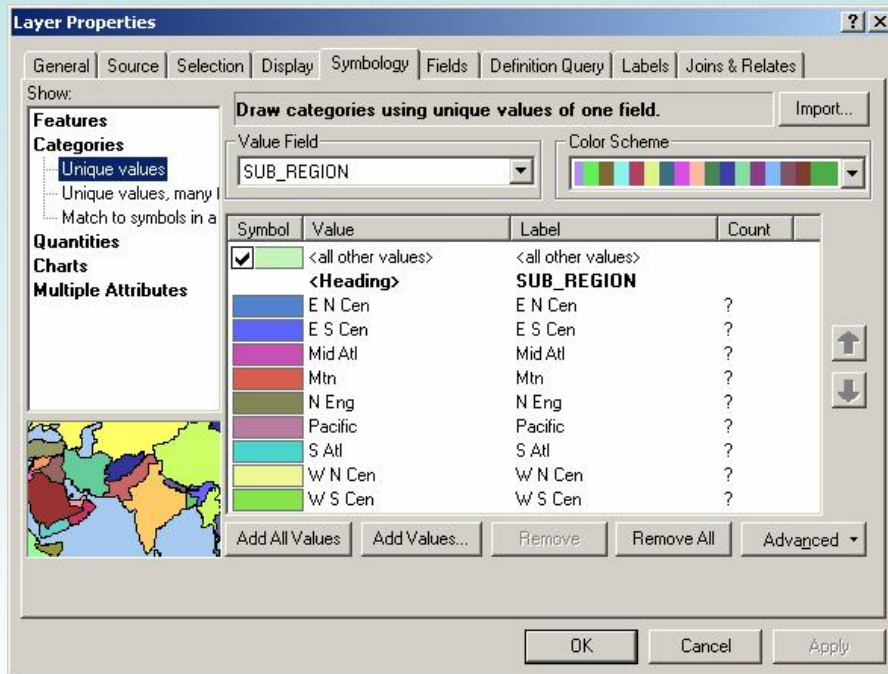
Creating a choropleth/thematic map: Continuous data

Map states of USA shaded by population density
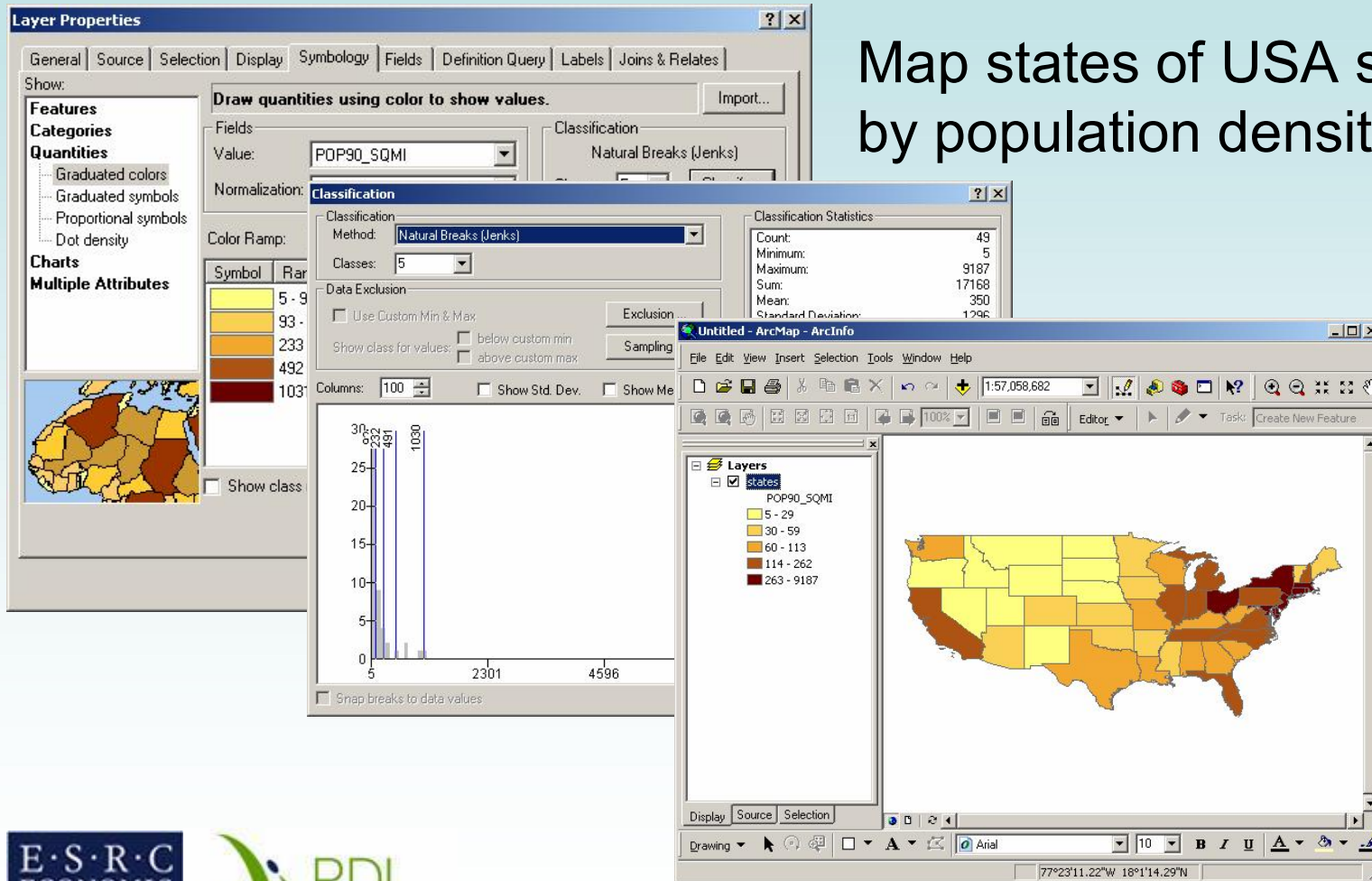
Really useful datasets/tools

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

Geo-Refer

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

Geo-Refer

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

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

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

RDI Researcher Development Initiative

www.neighbourhood.statistics.gov.uk

Geo-Refer

Other useful sites…

Royal Mail Postcode Finder and Address Finder

Registration required and limited number of uses

Digimap – Ordnance Survey mapping, inc. historical

JISC-funded subscription service at Edina

Google Earth and Microsoft Live Search Maps

Street mapping and aerial photography (and many others)

www.royalmail.co.uk

www.edina.ac.uk/digimap

www.maps.live.com

www.earth.google.com

Geo-Refer

Data linkage with GeoConvert
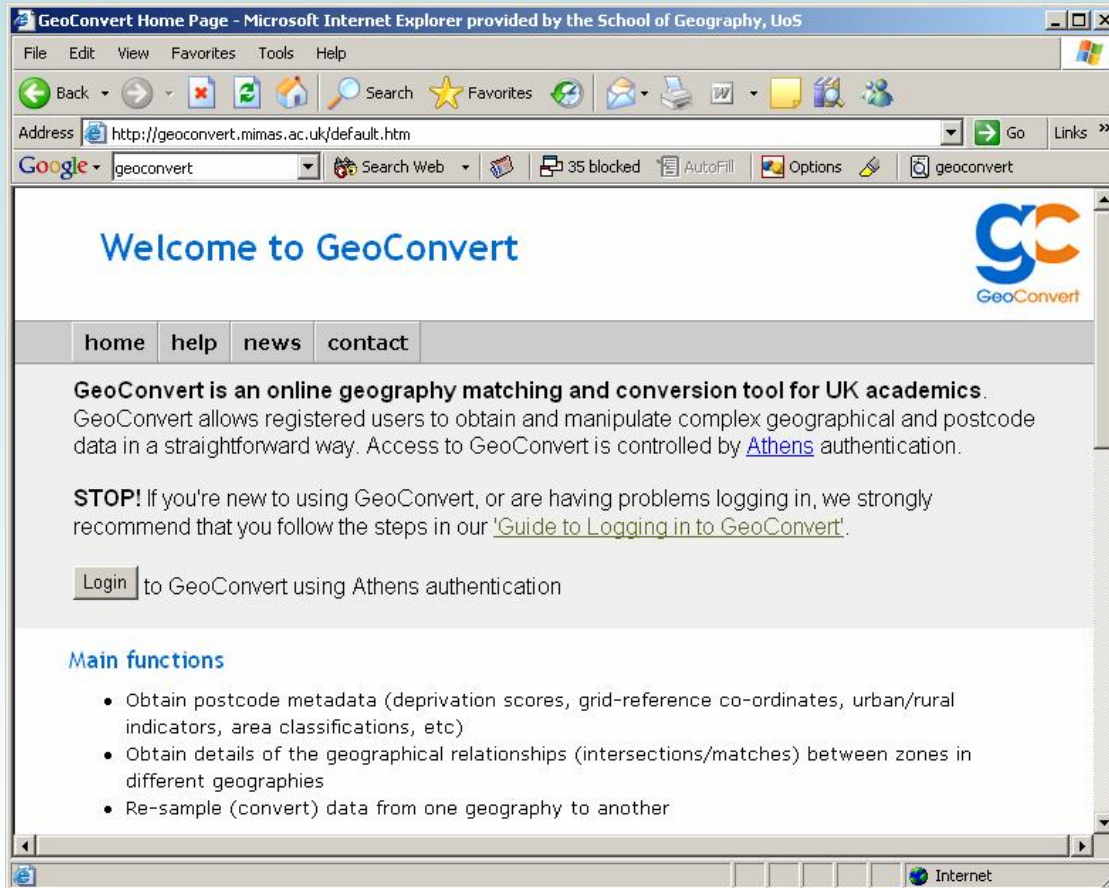
Developed by MIMAS

Online tools for NSPD extraction and matching

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

Geo-Refer

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!

RDI
Researcher
Development
Initiative

Geo-Refer

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

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

RDI Researcher Development Initiative

Geo-Refer

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

Geo-Refer

Lunchtime!

Geo-Refer

Practical workshop session

ESRC ECONOMIC & SOCIAL RESEARCH COUNCIL

RDI Researcher Development Initiative

Geo-Refer

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…

Geo-Refer

Presentation of examples

Workshop evaluation form

- Immediate feedback on workshop
- We would like to find our 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 Mimas online course booking form

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

Geo-Refer