Geographical referencing for social scientists: a practical workshop

30 May 2007

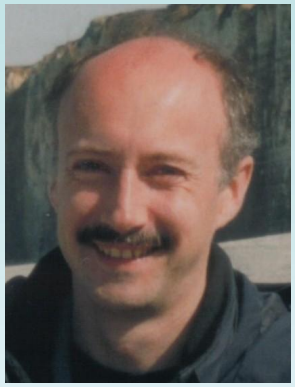
CCSR, University of Manchester

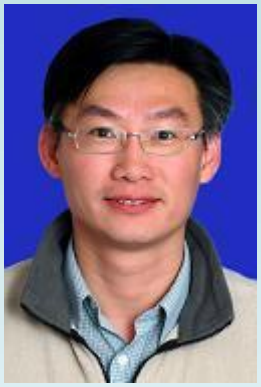
E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

RDI Researcher Development Initiative

Geo-Refer

Welcome and introductions





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

  www.geog.soton.ac.uk/geo-refer/workshop2.html

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

RDI Researcher Development Initiative

Geo-Refer

Housekeeping

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

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

RDI Researcher Development Initiative

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

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

RDI Researcher Development Initiative

Geo-Refer

Rules of engagement

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

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

RDI Researcher Development Initiative

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

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

RDI Researcher Development Initiative

Geo-Refer

Georeferencing examples

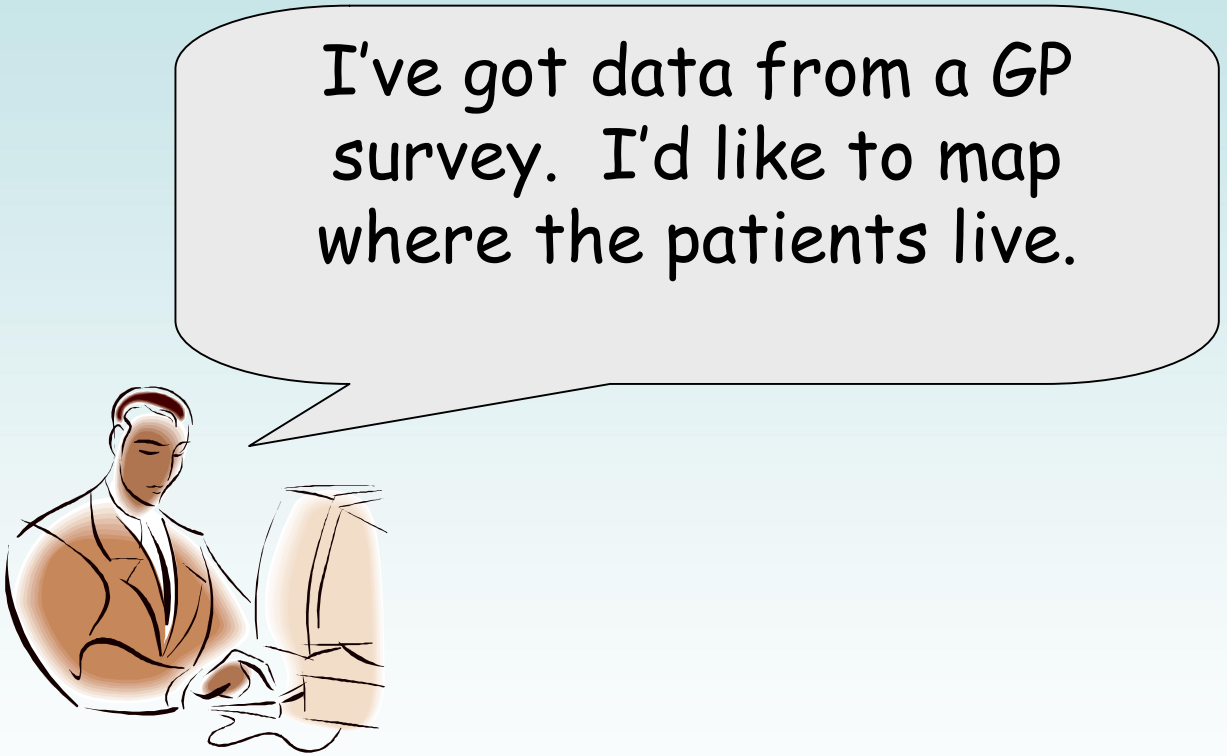
- Link survey results to census data
- Associate postcoded patient records to indices of deprivation
- Allocate questionnaire responses to urban/rural neighbourhood types
- Identify which service delivery locations fall within different policy areas
- Use GPS to record survey locations

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 …

Joe Bloggs

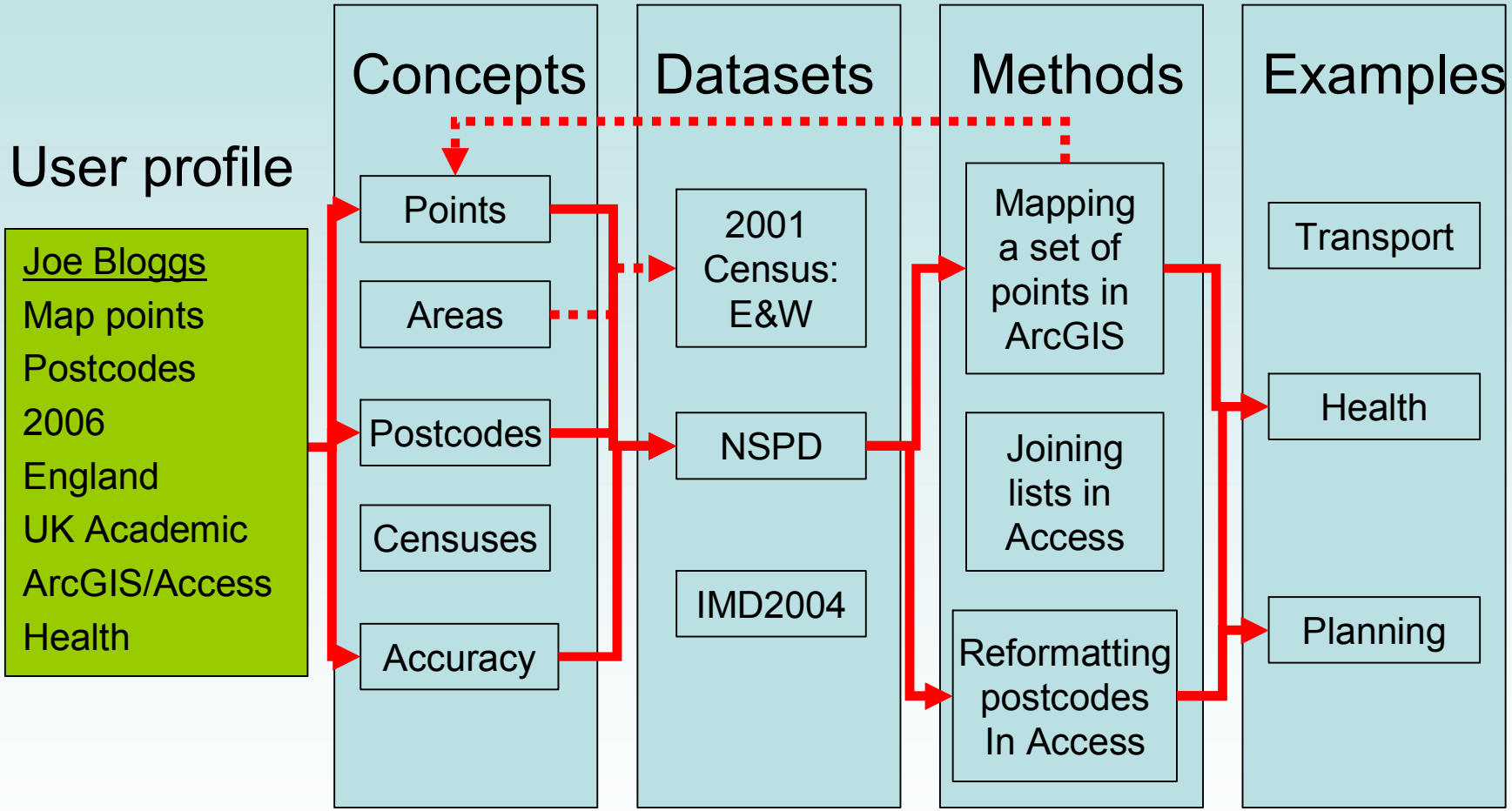


E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

RDI Researcher Development Initiative

Geo-Refer

Customised set of learning resources

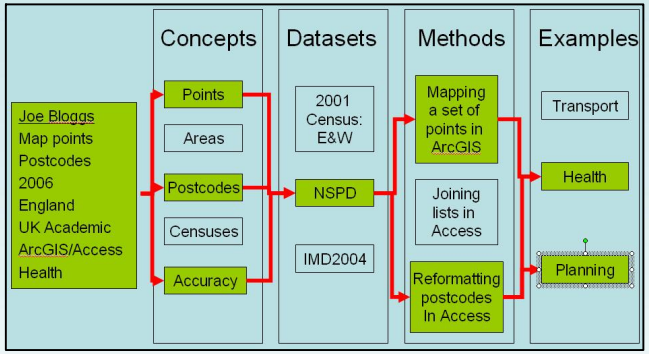


E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

RDI Researcher Development Initiative

Geo-Refer

Customised online tutorial



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

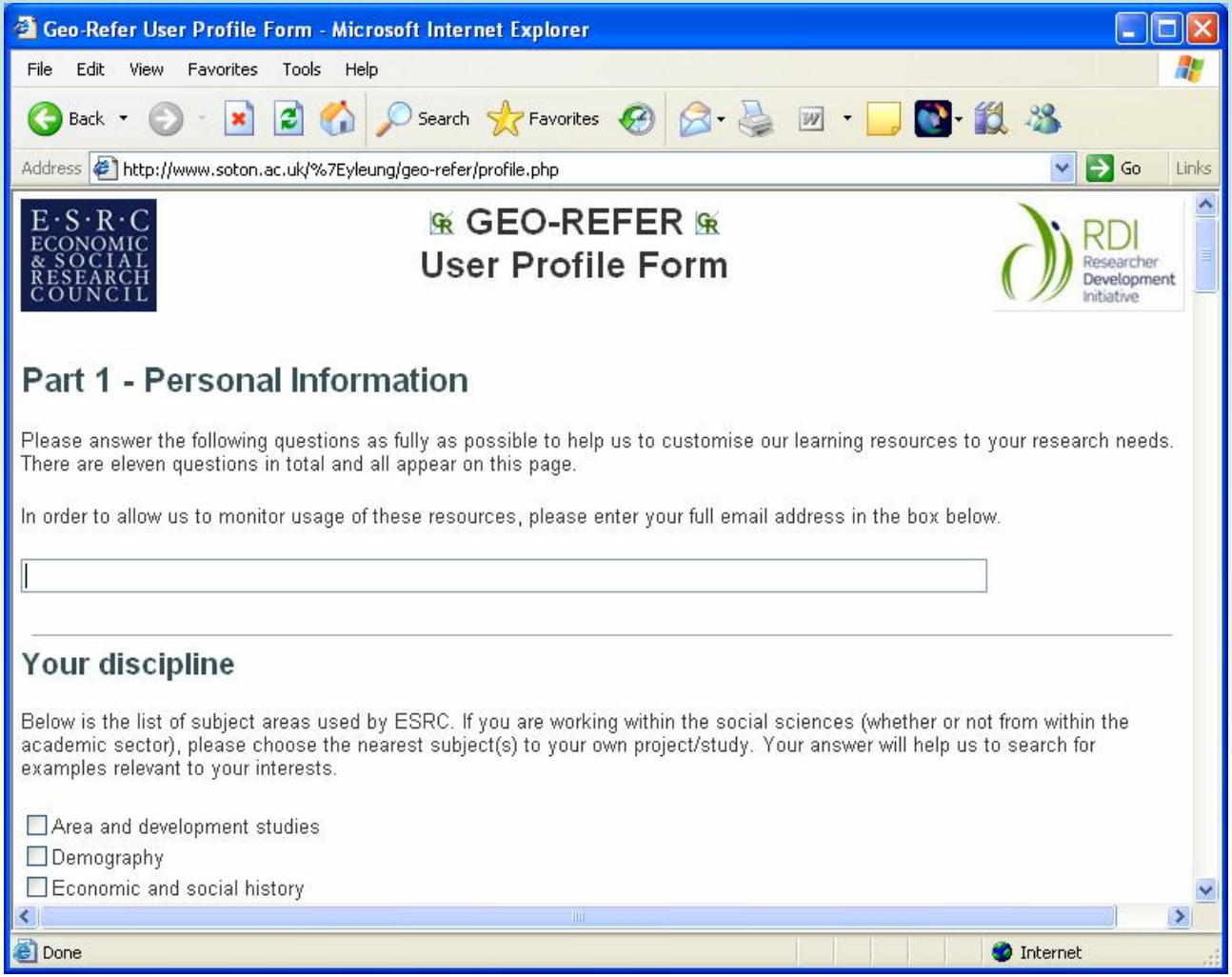
RDI Researcher Development Initiative

Geo-Refer

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!



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

RDI Researcher Development Initiative

Geo-Refer

Workshop programme

- 11.20 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

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

RDI Researcher Development Initiative

Geo-Refer

Georeferencing concepts

Geographical referencing of social science phenomena: 1

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

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

RDI Researcher Development Initiative

Geo-Refer

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

Photo: © David Martin

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

RDI Researcher Development Initiative

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

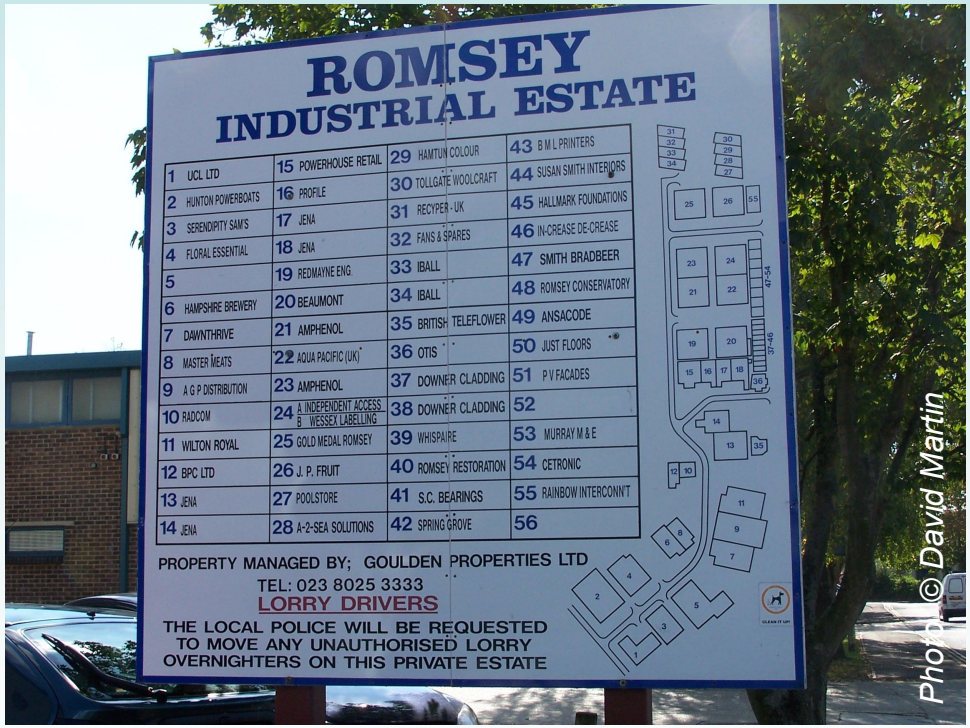
E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

RDI Researcher Development Initiative

Geo-Refer

Geographical referencing of social science phenomena: 5

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



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

RDI Researcher Development Initiative

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

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

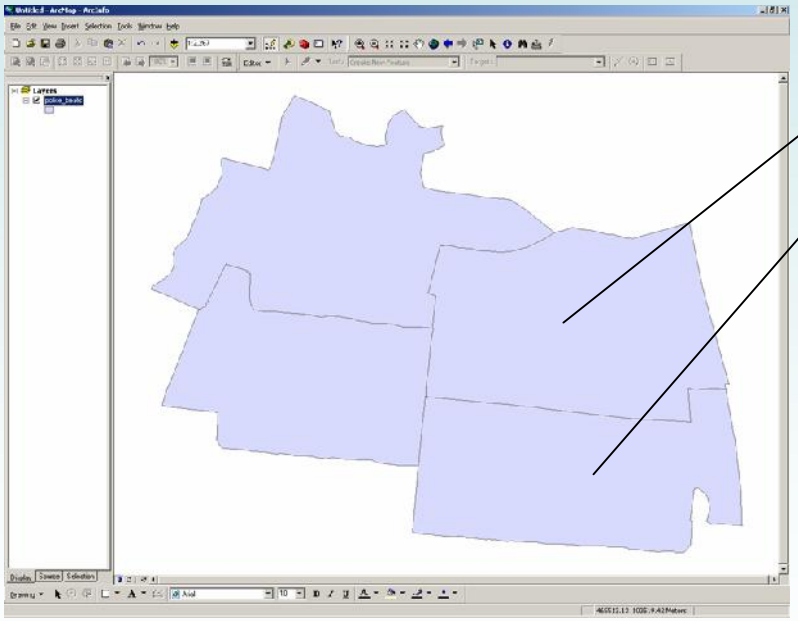
- Accurate location – free from locational bias (respondent lives in Manchester)
- Precise location – provides detail (respondent lives at 255b Stockport Road) but not necessarily accurate
- Need to assess and adopt different strategies according to purpose

Scale and projection

- Map scale
  - Land title plan: 1:1,250
  - Explorer map: 1:25,000
- 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 ou

- Attributes: values, characteristics

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

RDI Researcher Development Initiative

Geo-Refer

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!

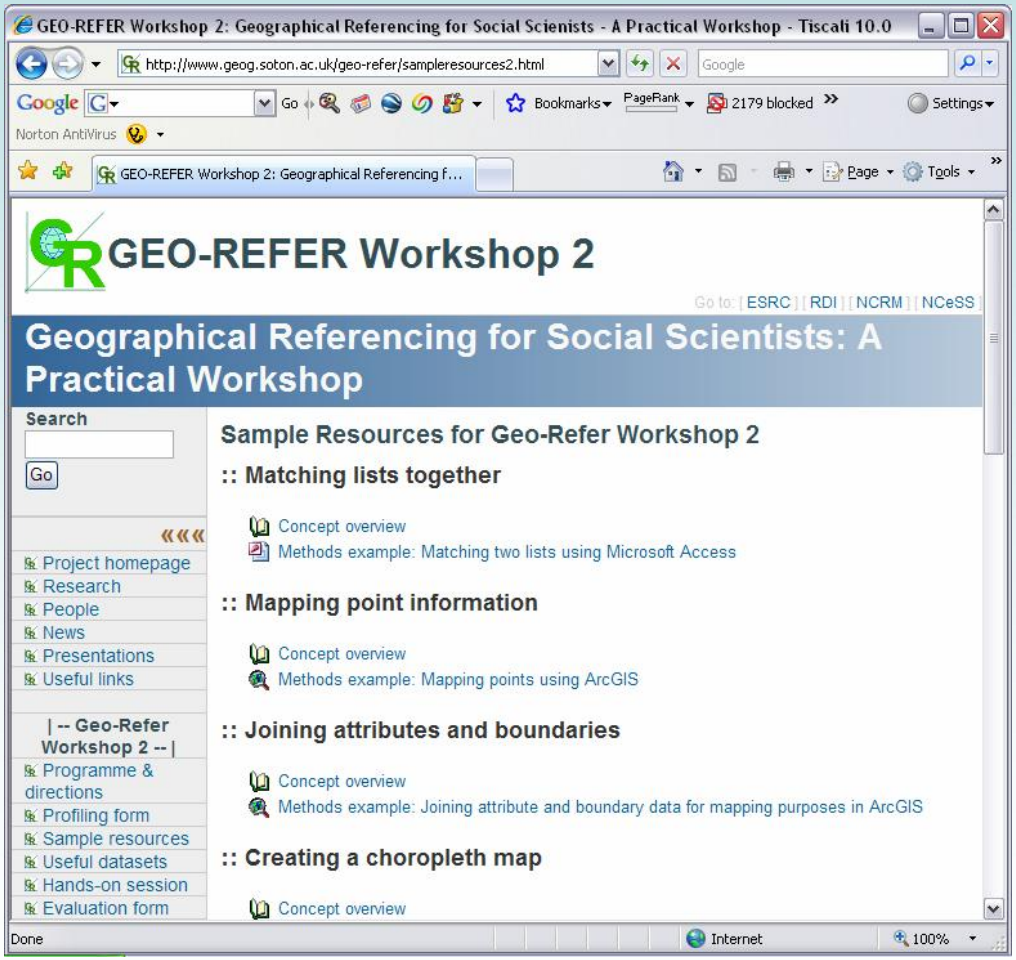
What about geographical information systems (GIS)?

- 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
- Workshop website:
  - Sample resources
- This afternoon:
  - Try them out!



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

RDI Researcher Development Initiative

Geo-Refer

Georeferencing methods

1 Data linkage
  - Tabular
  - Spatial
    - Allocation and aggregation

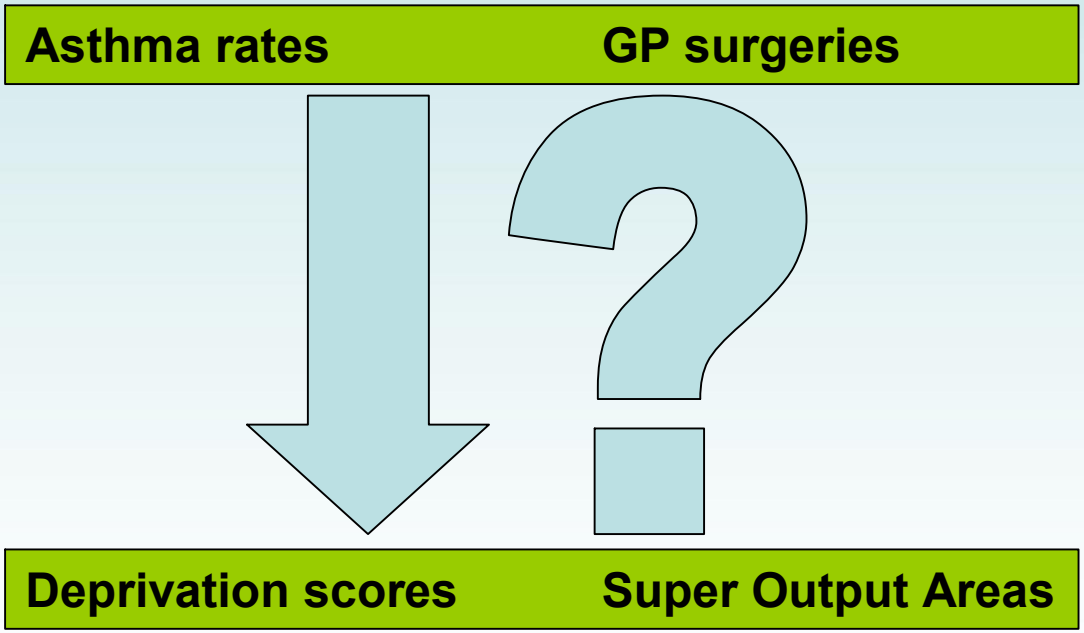
2 Mapping
  - Points
  - Areas
    - Choropleth/thematic maps

Methods Part 1: Data linkage

- Link datasets through geography
- Link may be:
  - Tabular, using common items in dataset
  - Spatial, using topology/geographical features

Tabular data linkage example

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



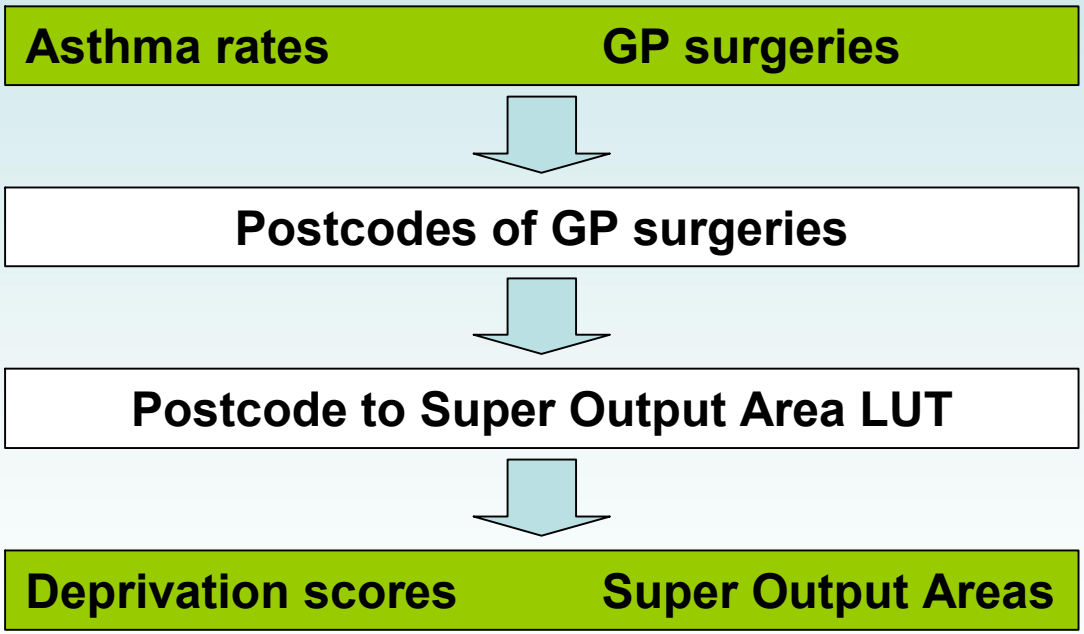
E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

RDI Researcher Development Initiative

Geo-Refer

Tabular data linkage example

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



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

RDI Researcher Development Initiative

Geo-Refer

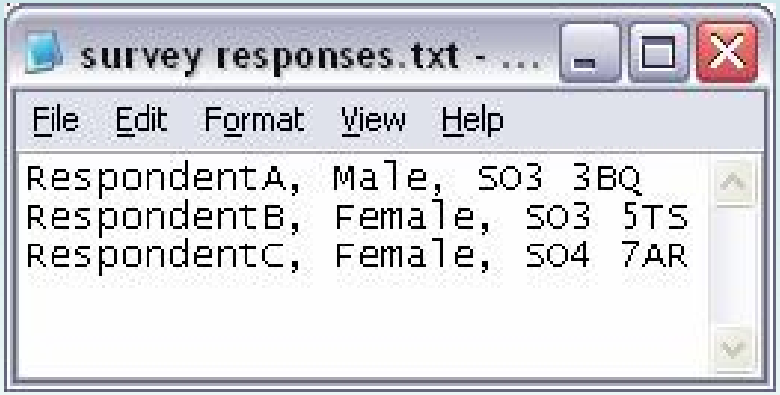
Tabular linkage tools

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

- GeoConvert – online tool specific to recent UK postcode-based lookup (developed from Convert, about to be released)

Data linkage example: input tables

Coded survey responses



Respondent ID

Gender

Postcode

Postcode directory

Postcode

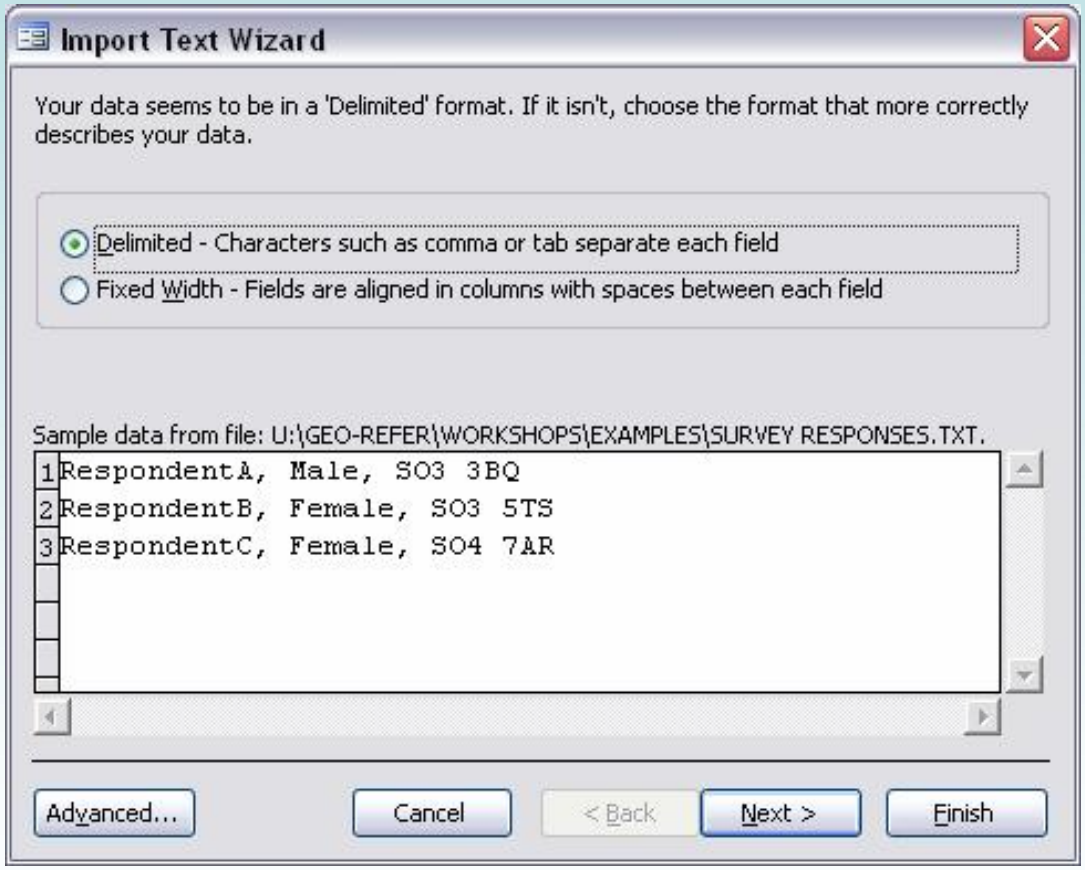
Ward code

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

RDI Researcher Development Initiative

Geo-Refer

Import external data files



Get external data > import

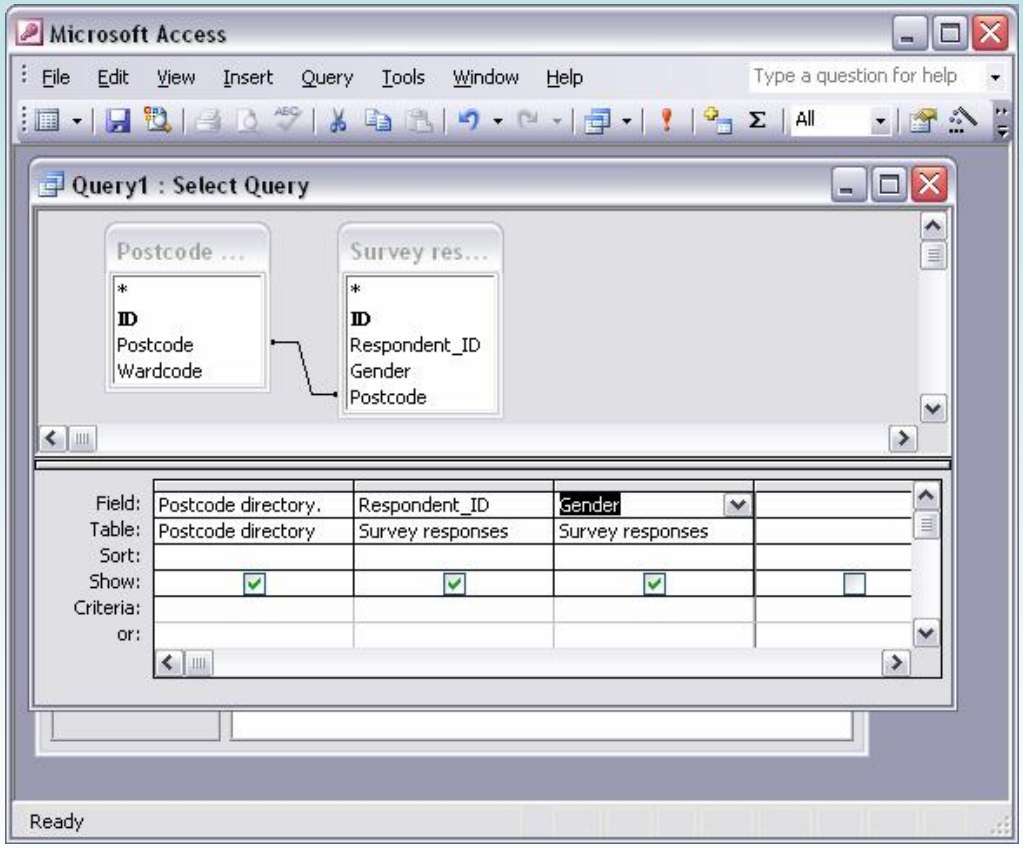
We need to describe the data structure (here comma separated)

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

RDI Researcher Development Initiative

Geo-Refer

Build new Access query



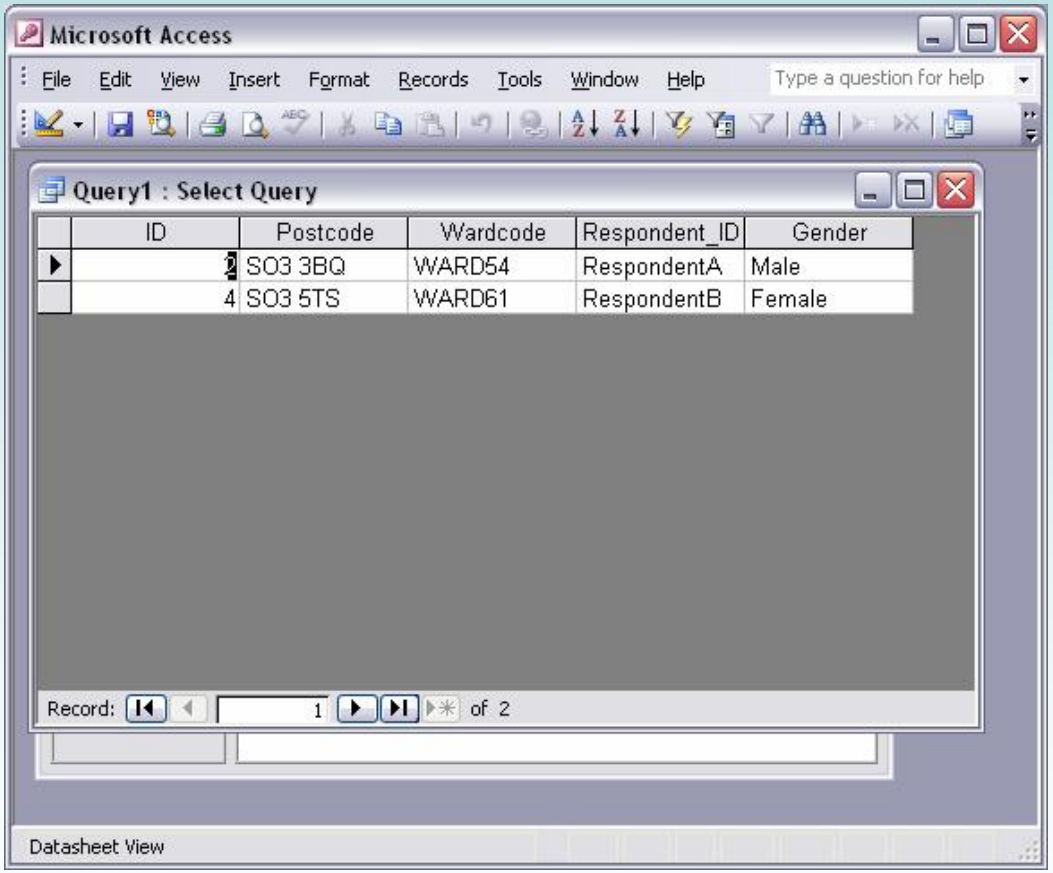
We now add both tables to a new query, link them by the postcode field and select the items required in output

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

RDI Researcher Development Initiative

Geo-Refer

Matched lists



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

NB not all are matched!

Why??

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

RDI Researcher Development Initiative

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!

Georeferencing methods

1 Data linkage
- – Tabular
- – Spatial
  - Allocation and aggregation

2 Mapping
- – Points
- – Areas
  - Choropleth/thematic maps

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

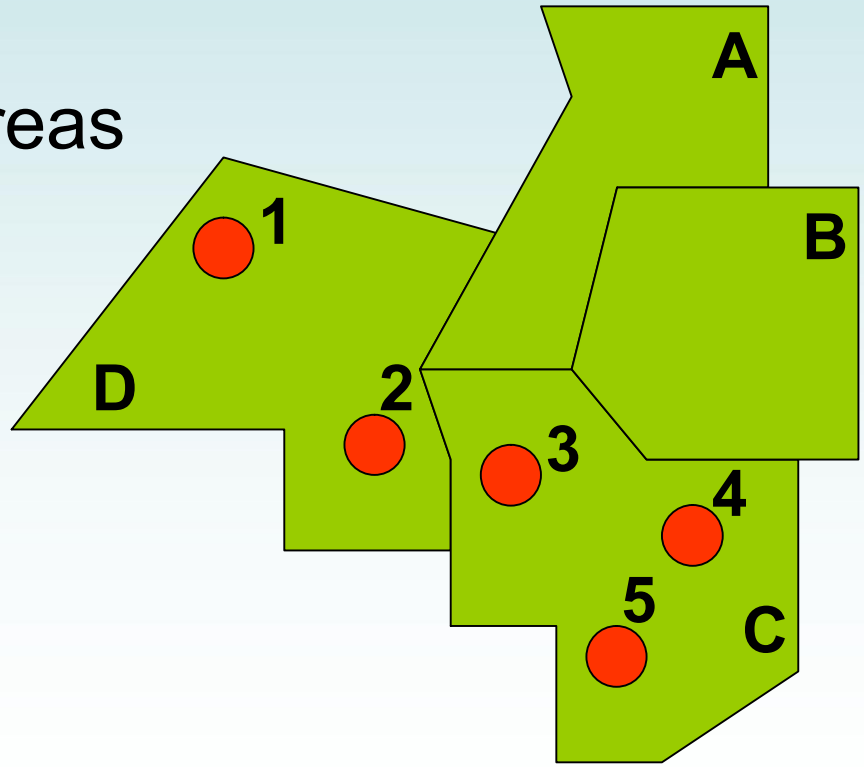
Car crimes: occur at many locations which cannot be addressed or postcoded

Spatial data linkage example

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

Police beat areas

Car crimes



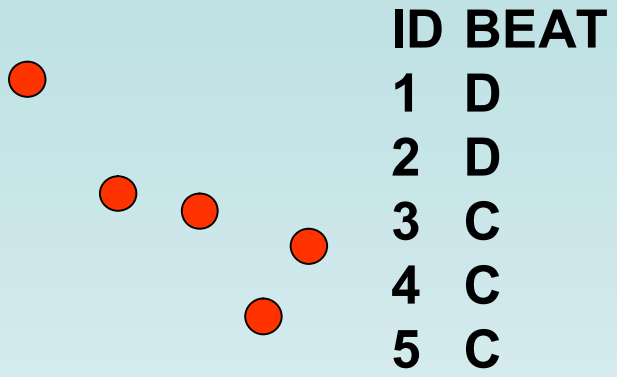
E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

RDI Researcher Development Initiative

Geo-Refer

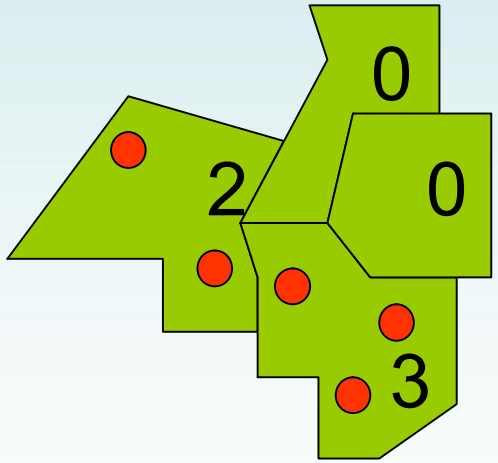
Allocation and aggregation

- Allocate
  - E.g. allocate attributes of police beats to car crime locations



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

- Aggregate
  - E.g. aggregate number of car crimes in each police beat



- Same spatial linkage operation, but different outputs

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

RDI Researcher Development Initiative

Geo-Refer

Spatial linkage tools

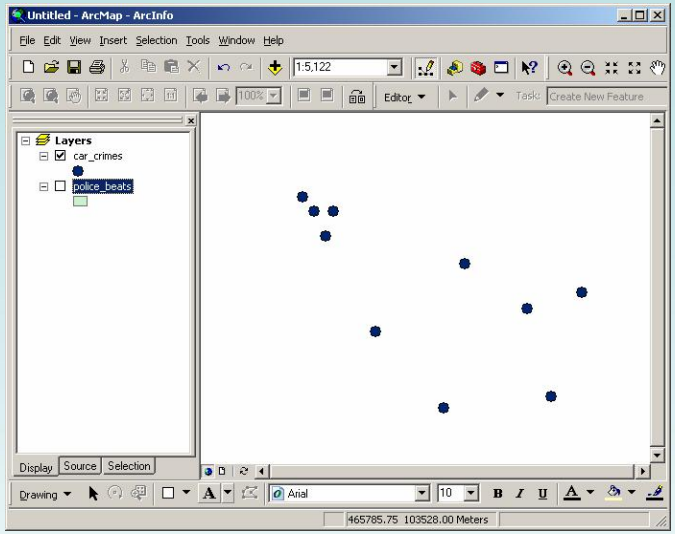
- Spatial linkage usually done using a GIS
  - Various GIS software packages
  - ESRI (ArcGIS, ArcInfo, ArcView)
  - MapInfo
  - IDRISI

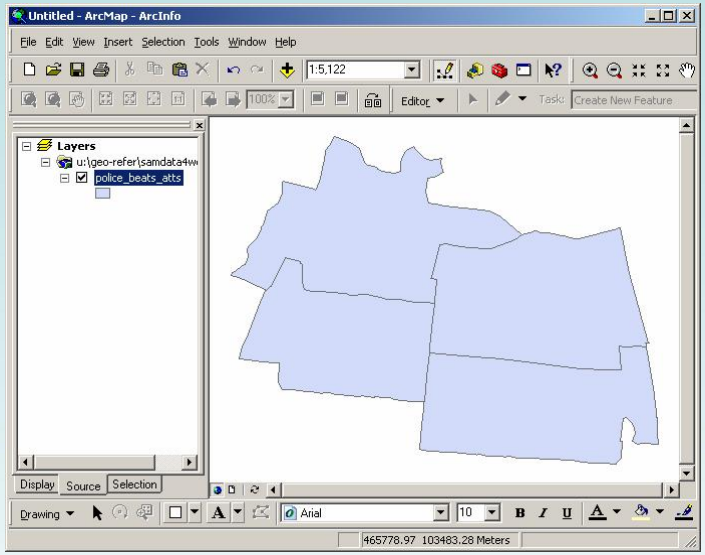
- Remember: This is not a GIS course!
- Current worked examples use ArcGIS

Allocation: joining poly data to points data

Car crimes: points

Police beats: areas





Spatial link?

Point-in-polygon

Attributes of CRIMES

OID	CRIMEID	X	Y
0	1	465474	103429
1	2	465490	103409
2	3	465517	103409
3	4	465507	103374
4	5	465703	103334
5	6	465791	103271
6	7	465869	103294
7	8	465577	103238
8	9	465825	103146
9	10	465673	103129

Record: 1 Show: All Selected

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

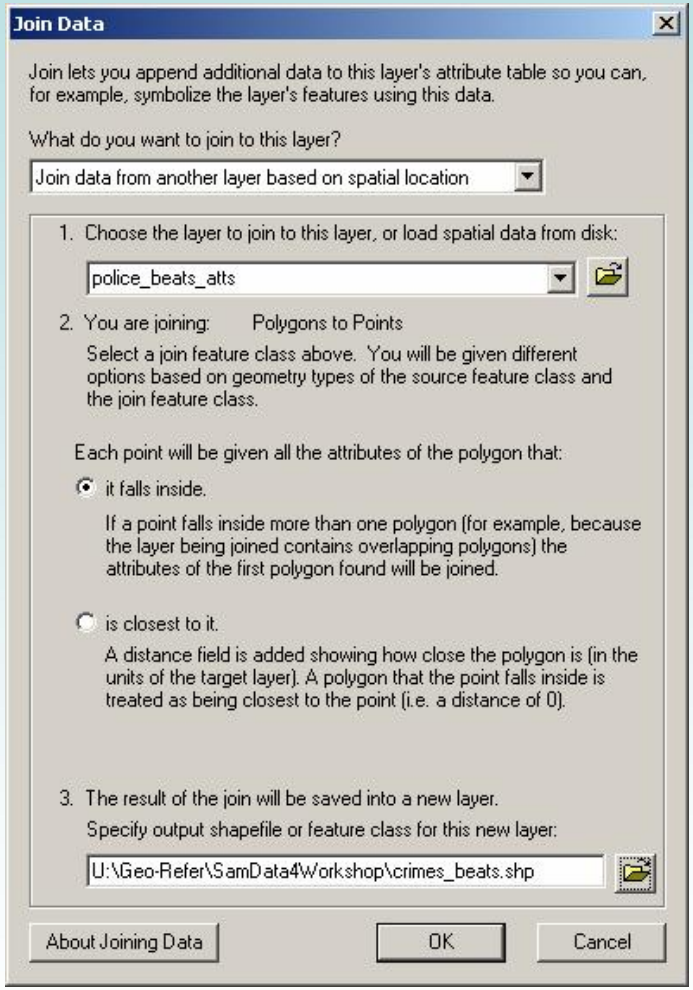
Allocate polygon attributes to points

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

RDI Researcher Development Initiative

Geo-Refer

Allocation: joining poly data to points data



Spatial join: polygons to points

Sample output

Attributes of crimes_beats

FID	Shape	FI	CRIMEID	X	Y	FI	BEATID	NAME	NOFFS	AUTH
0	Point	7	8	465577	103238	0	1	Westend	30	Anywhere
1	Point	8	9	465825	103146	1	2	Southend	20	Anywhere
2	Point	9	10	465673	103129	1	2	Southend	20	Anywhere
3	Point	4	5	465703	103334	2	3	Eastend	5	Nowhere
4	Point	5	6	465791	103271	2	3	Eastend	5	Nowhere
5	Point	6	7	465869	103294	2	3	Eastend	5	Nowhere
6	Point	0	1	465474	103429	3	4	Northend	10	Nowhere
7	Point	1	2	465490	103409	3	4	Northend	10	Nowhere
8	Point	2	3	465517	103409	3	4	Northend	10	Nowhere
9	Point	3	4	465507	103374	3	4	Northend	10	Nowhere

Record: 1 Show: All Selected Records (0 out of 10 Selected.) Option

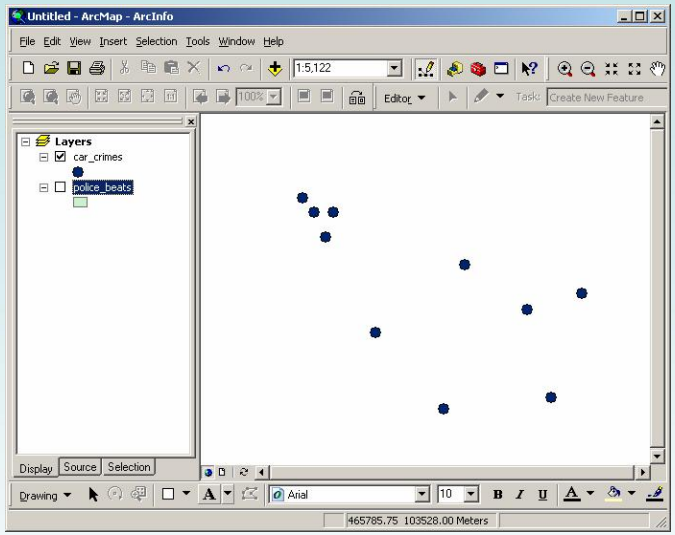
E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

RDI Researcher Development Initiative

Geo-Refer

Aggregation: joining points data to poly data

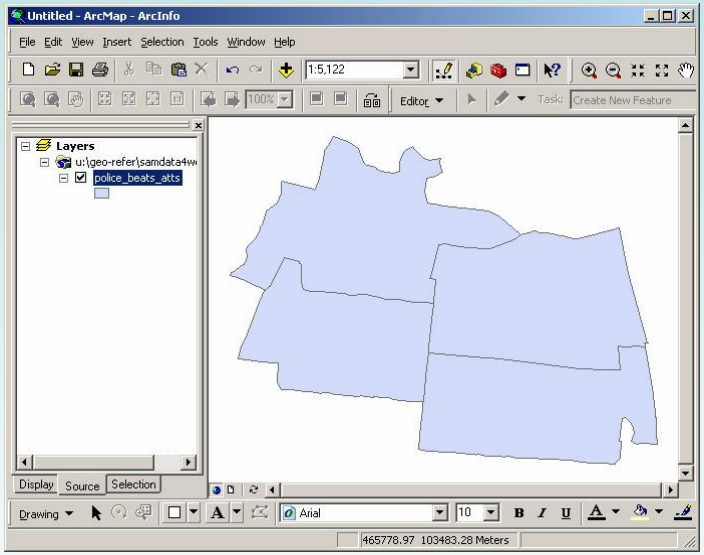
Car crimes: points



Spatial link?

Point-in-polygon

Police beats: areas



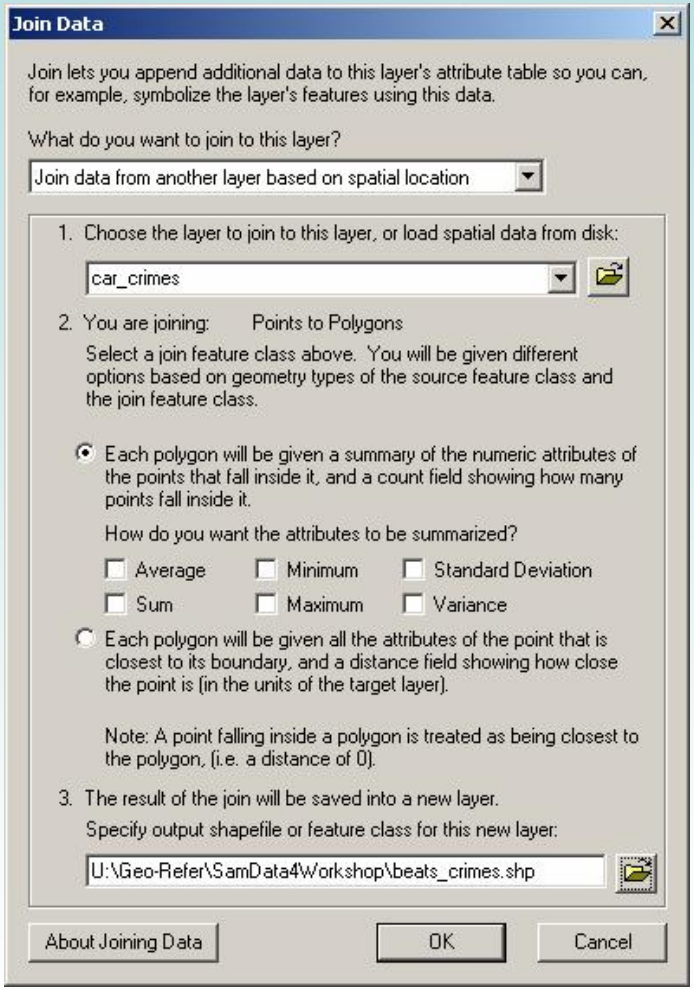
Aggregate points and/or their attribute data to polygons

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

RDI Researcher Development Initiative

Geo-Refer

Aggregation: joining points data to poly data



Spatial join: points to polygons

Sample output

Attributes of beats_crimes

FID	Shape	FID_1	BEATID	Count_
0	Polygon	0	1	1
1	Polygon	1	2	2
2	Polygon	2	3	3
3	Polygon	3	4	4

Record: 1 Show: All Selected Records ([

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

RDI Researcher Development Initiative

Geo-Refer

Georeferencing methods

1 Data linkage
- Tabular
- Spatial
  - Allocation and aggregation

2 Mapping
- Points
- Areas
  - Choropleth/thematic maps

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)

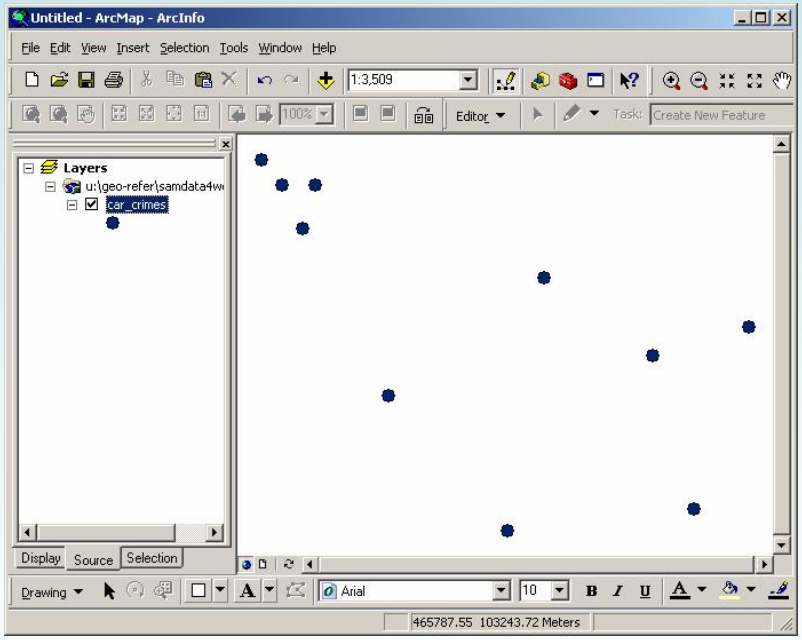
- Map attributes of areas
  - E.g. Link survey responses on attitudes to recycling to LADs then map response rates by LADs

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?



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

RDI Researcher Development Initiative

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)

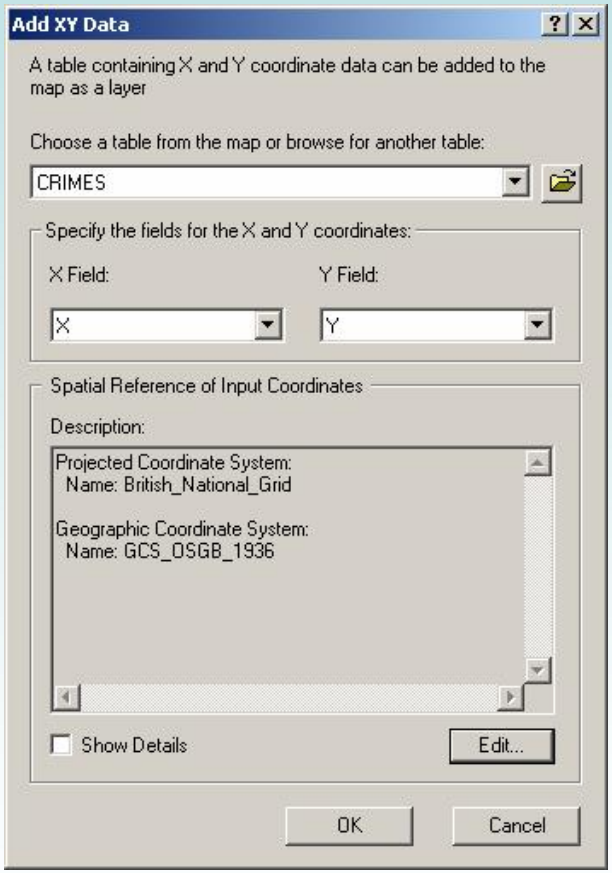
Import into Access and define fields e.g. ID, X, Y

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



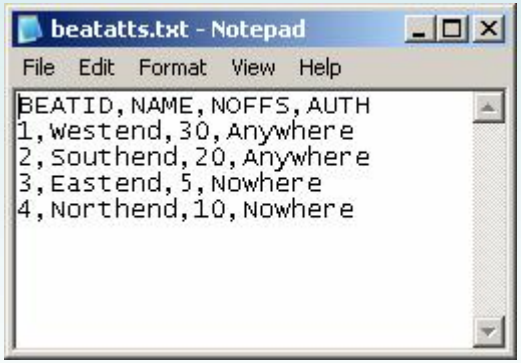
E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

RDI Researcher Development Initiative

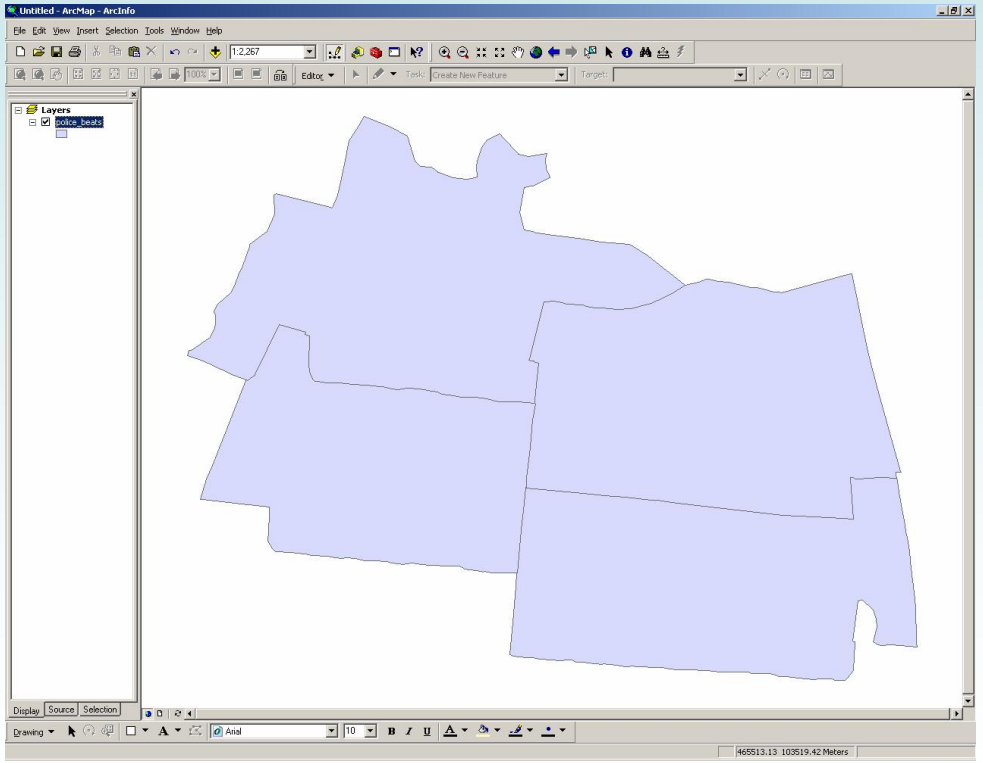
Geo-Refer

Linking attribute data to boundary data

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



?



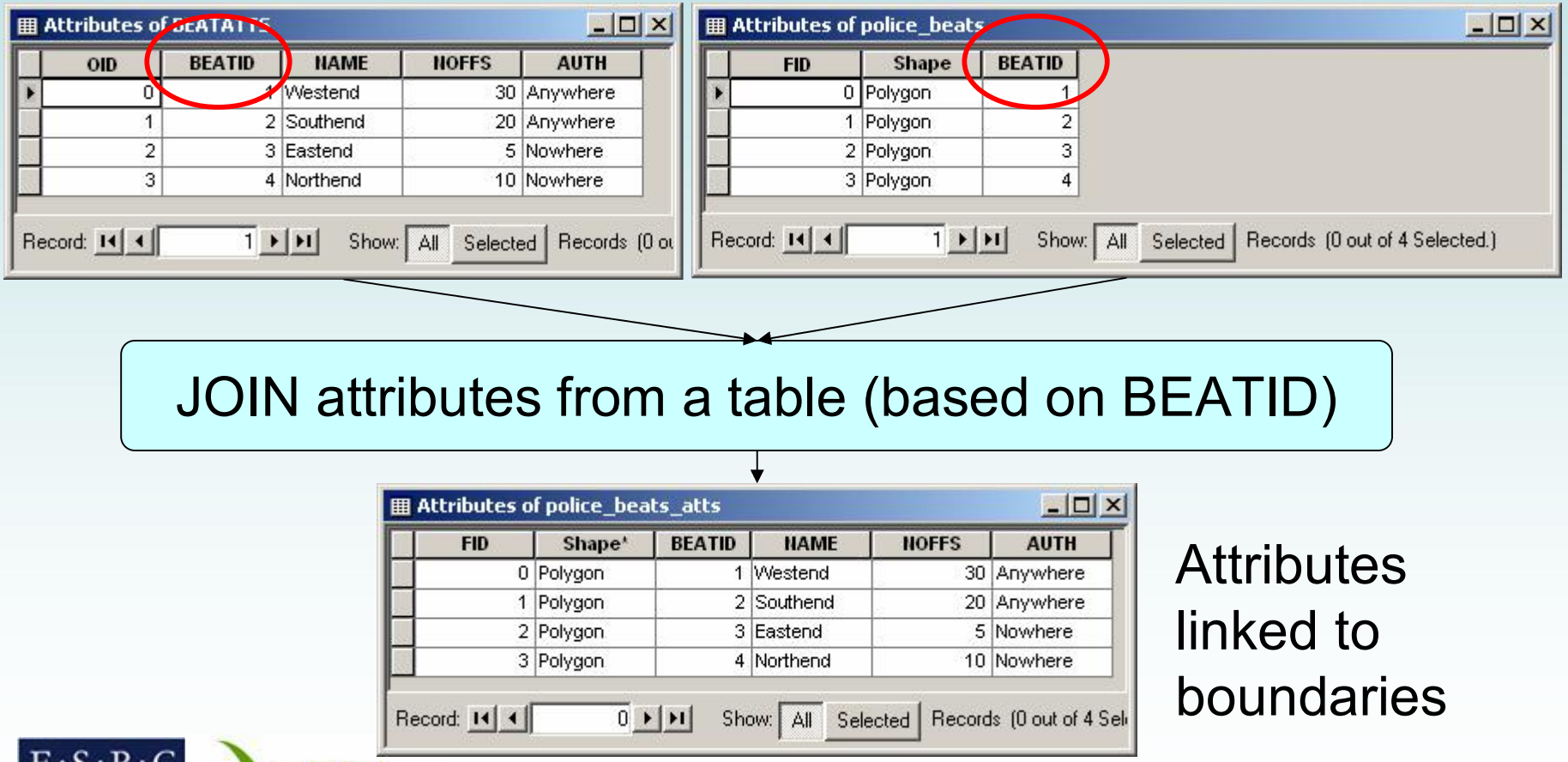
E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

RDI Researcher Development Initiative

Geo-Refer

Linking attribute data to boundary data

Define common field in attribute table and in shapefile attribute table



Attributes linked to boundaries

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

RDI Researcher Development Initiative

Geo-Refer

Creating a choropleth/thematic map

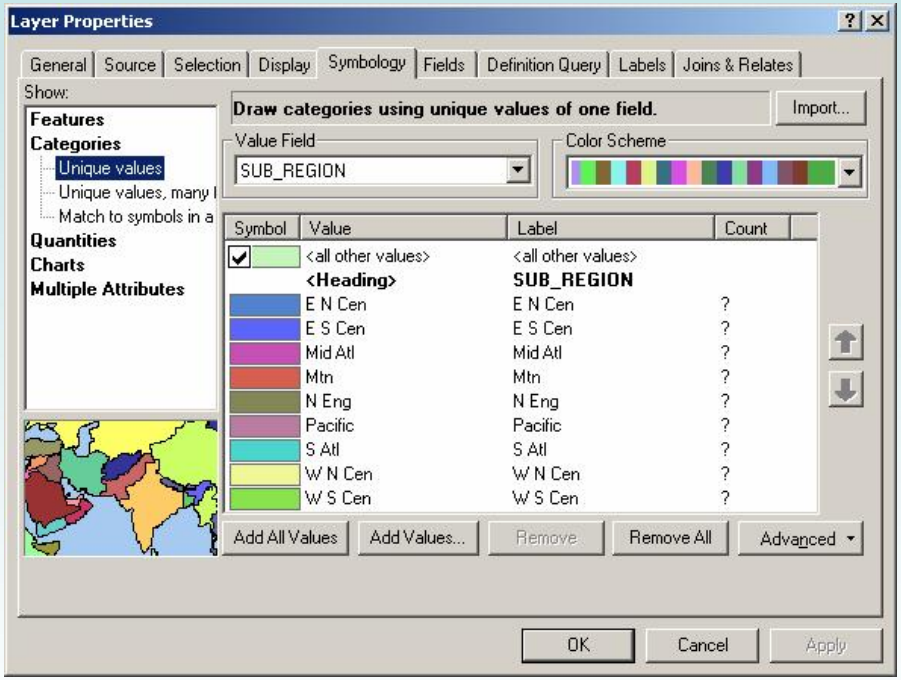
- OK, I've got my attributes linked to boundaries, but how do I map them?

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

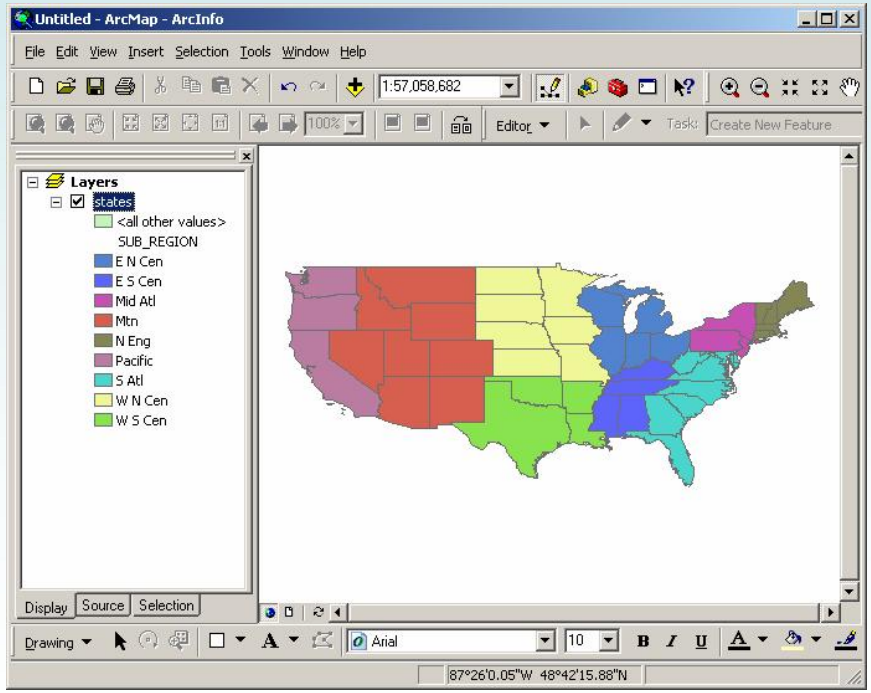
Creating a choropleth/thematic map

- 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

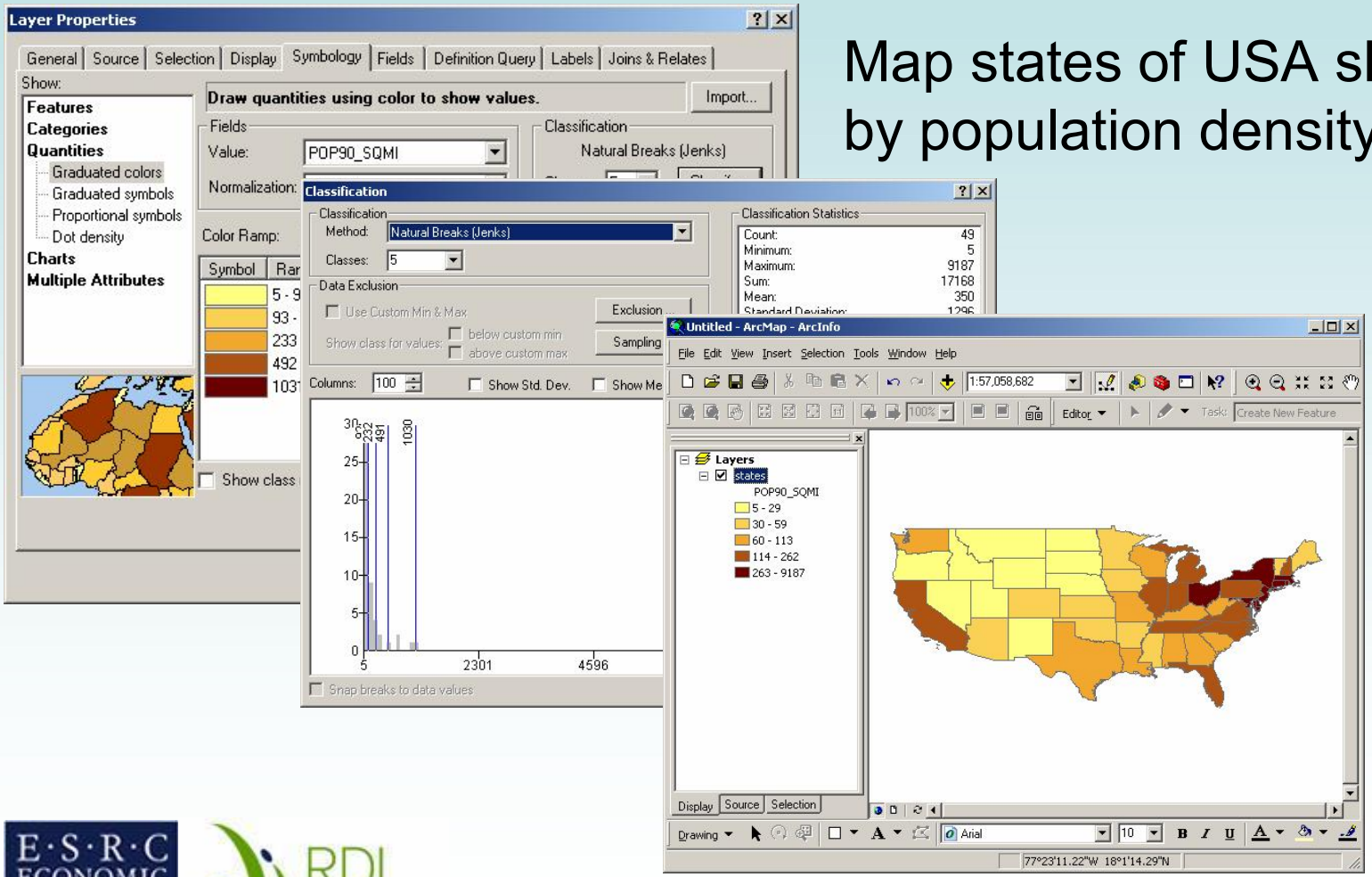


E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

RDI Researcher Development Initiative

Geo-Refer

Creating a choropleth/thematic map: Continuous data



Map states of USA shaded by population density

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

RDI Researcher Development Initiative

Geo-Refer

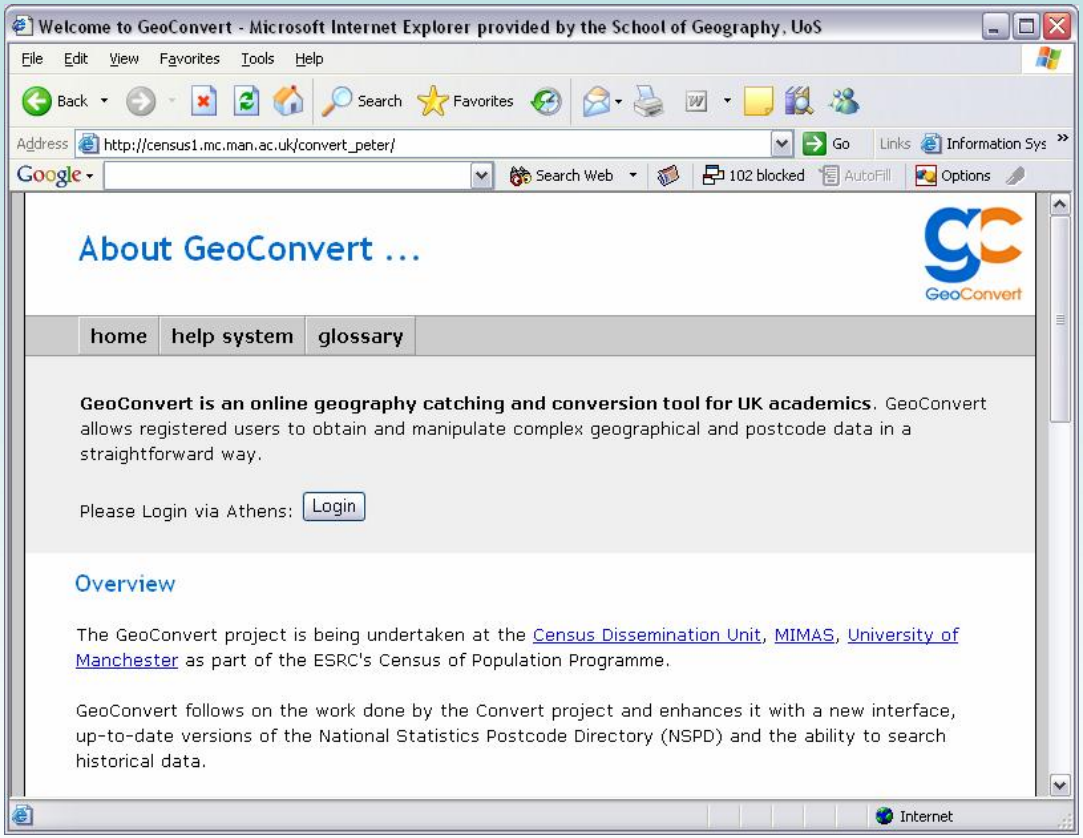
Really useful datasets

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

Data linkage with GeoConvert



Developed by MIMAS

Online tools for NSPD extraction and matching

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

RDI Researcher Development Initiative

Geo-Refer

GeoConvert functions

- Obtain postcode-related information
- Identify spatial relationships between geographies
- Convert data between geographies
- Obtain a subset of rows from a postcode directory, together with supplementary information

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

Using Geo-Refer online resources

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

RDI Researcher Development Initiative

Geo-Refer

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!

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

RDI Researcher Development Initiative

Geo-Refer

Practical workshop session

E·S·R·C 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…

Presentation of examples

Emerging findings

Emerging findings…

- GeoConvert is a very useful tool!
  - But it is not quite finished…
- NSPD is a very big file!
  - Subsetting?
- Incomplete postcodes such as sectors (e.g. SO51 6) can be treated as areas
- Cannot make 1:1 matches of incompatible geographies (esp. at similar scales!)

- ONS output area classification – Sheffield/Casweb has names, ONS only has codes
- Drive times? (crow fly distance > drive time surface models > full network models) increasing GIS complexity, but all possible
- MS Access for non-Athens use: lots of control over formats etc. but more work

. . .

- Digimap www.edina.ac.uk/digimap - access to OS digital data for subscribing academic institutions (loc. govt. MSA)
- Technical possibility does not equate to common sense

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!

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

RDI Researcher Development Initiative

Geo-Refer