Counting Cows and Cabbages – Web-based extraction, delivery and discovery of geo-referenced data

Stuart Macdonald

EDINA National Data Centre

University of Edinburgh, Scotland

EDINA context

- A JISC-Funded National Data Centre for Tertiary Education set up in 1995

  - Based at the University of Edinburgh, Scotland

  - Offers the UK tertiary education and research community networked access to a library of data, information and research resources

EDINA context cont'd..

- Range of bibliographic resources
- Multimedia and image services

- Geo-spatial data and geo-referenced information:
    - **UKBORDERS** digitised boundary outlines
    - **Digimap** online OS mapping products
    - **agcensus** – online Agricultural census data

- Geo-related development projects - Go-Geo!, geoXwalk, e-MapScholar…

Agricultural Census background

- Conducted in June of each year in the UK since 1866

- Provides information on distribution and extent of crop & horticultural production and rearing of livestock.

- Indicator of regional change in agricultural activity over time.

Data collection

- Each farmer declares the agricultural activity on the land in June each year via a postal questionnaire

- DEFRA, SEERAD (& Welsh Assembly since 1999) collect > 150 items of data and publish information relating to farm holdings for recognised geographies

Problems with published data

- Data potentially disclosive

- Parishes vary considerably in shape & size

- Landuse & capability vary across parish

- Under- & over-reporting of agriculture in parish

- Livestock may be elsewhere

- "Snapshot" of activity on 1st June

Parish to grid

- Conversion algorithms developed by EUDL for grid square estimates

- Parish Framework defines each parish in terms of 1km grid squares

- Land-use Framework classifies 1km grid squares to land-use category

- 1km data still confidential – aggregation to 2km, 5km & 10km grid square estimates

Data uses

- Epidemiology / Public health
  - Pesticide effects on congenital 'small eye' syndrome

- Market research

- GIS modelling

- Environmental
  - Impact of sheep on heather coverage
  - Wind farm placement

Existing service....

- No online access

- Algorithms written in main frame era

- Fetching & delivering of data
  - Command-driven
  - 'manual'
  - reference interview via telephone/email
  - time consuming

- User expectations not matched

New EDINA agcensus

- EDINA online agcensus service offers agricultural census data to 3 client communities:

- Academic -   subscription

- Commercial – price on request

- Policy makers – price on request
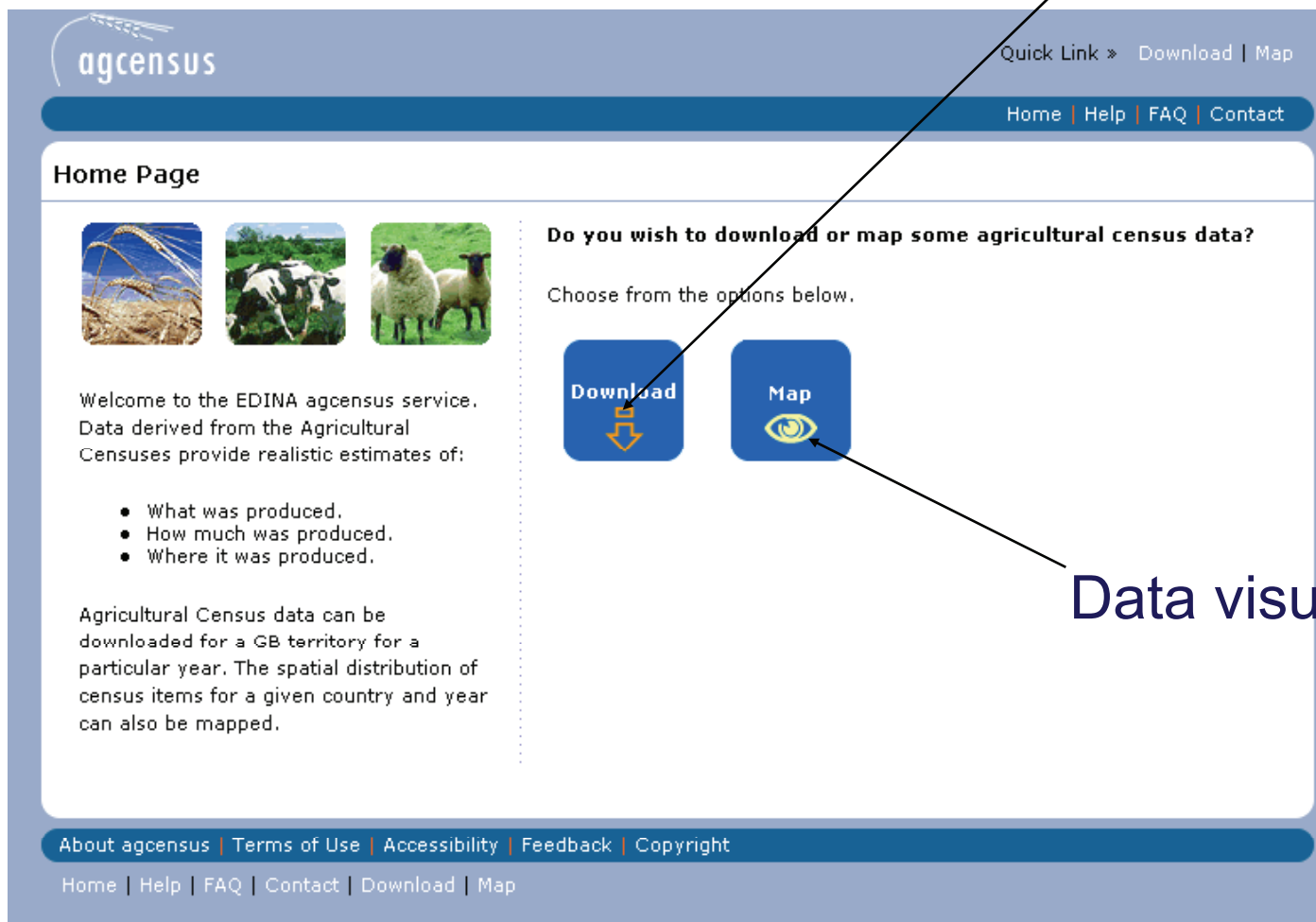
Technology

- Java-based interface and data processing algorithms

- Backend data stored in MYSQL database

- Data visualisation uses image generation Servlets

Problems encountered

- Alternate visualisation for maps (area selection)

- Accessibility issues
  - allowing both javascript & non-javascript equivilant functionality

- Fast access to data
  - storage versus 'on the fly' delivery

Data download

Data visualisation

7 Step data access procedure:

- Select country (Sco, Eng, Wal)
- Select year
- Select census item(s)
- Select grid size: 2km, 5km, 10km
- Selecting extent of data coverage
- Data selection summary
- Download or visualise selected data
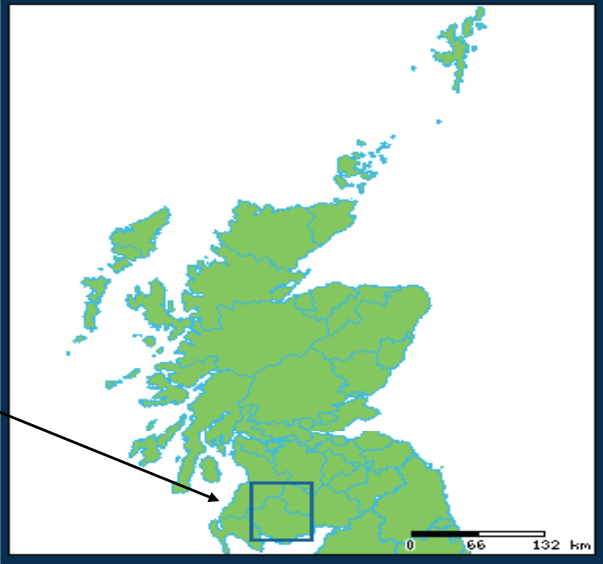
Selecting extent of data coverage

Extent tool for small area analysis

Step 7. Visualisation of data distribution at chosen resolution

Right-click to save image file
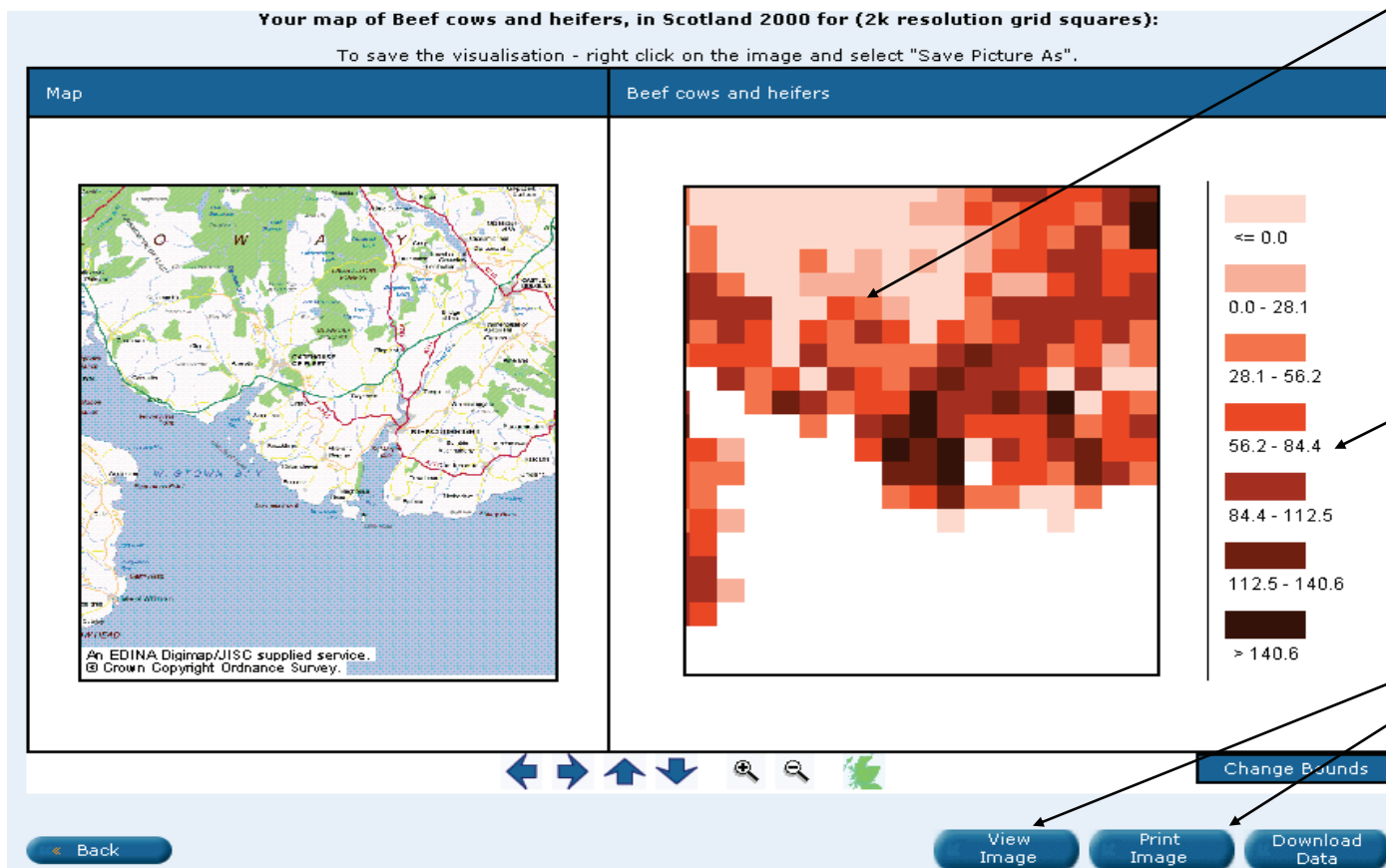
User defined legend

View or print image

Service issues

- Authentication   (ATHENS)
- Documentation   (QRGs, poster, handouts)
- Field Trialing   (usability, feedback)
- Help/Metadata
- Subscription model
- Training/outreach
- User support

Future developments

- Visualising change over time
- Statistical reporting
- Ability to add 2 or more items together
- Teaching dataset
- Other visualisations (e.g. histograms, pie charts etc)

The bigger picture …

How can UK academic geo-data users:

- find out about the **agcensus** data?
- find out what other geo-data exists within (and outwith) UK academia?
- ascertain quality and suitability?
- access data & find out how to use it?

Go-Geo! – Geographic Information Portal

- JISC-funded project - EDINA & UK Data Archive as partners

- to promote greater awareness of geo-spatial data created by UK academia within and without the wider UK GI community
  - part of the **UK GIgateway service** hosted at EDINA

- Goes beyond discovery to provide access to data

Metadata

- Central to such a service is metadata describing geo- referenced data

- Metadata originally used was modified NGDF standard
  - moved to ISO 19115 March 2003 which maps to Dublin Core

Metadata creator - Online Java-based metadata tool

Resources

Simple search

Advanced search

Advanced searching

- Data type
- Location
- Text
- Date range
- Summary
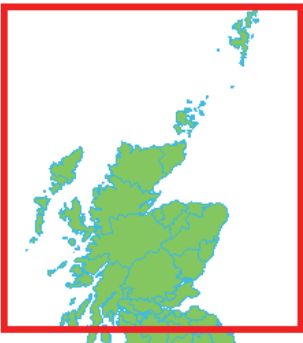
Map browser

Sample record

Geo-X-Walk gazetteer project

- Z39.50 compliant digital GB & Irish gazetteer service
  - Support geo-searching within the JISC-IE
  - Provide researchers & teachers with online reference tool

- Gazetteer will contain more than just placenames, grid references etc – will also classify features into types:
  - Settlements as areas, rivers as lines etc
  - Stores geo-spatial 'footprint' against each feature
  - Geo-searching transparent by 'cross-walking' different geography types – introducing proximity/distance into search strategies

Geo-X-Walk cont'd..

- Geo-parser – software that 'reads' a document/resource that contains placenames & automatically identifies their occurrence

- Identified placename can be compared against gazetteer entries which means access to 'alternate' geographies
  - Provides means to assign geo-tag (e.g. grid ref.) to implicit geo-referenced material (e.g. placename)

- Combination of parser & digital gazetteer has potential for powerful geographic-based searching across a range of otherwise disparate resources

E-MapScholar

- A learning and teaching project to develop tools that enhance and support the use of geo-spatial data

  - Case studies
  - Online L&T materials
  - Teaching content management system
  - Virtual work placement

Summary

- The new EDINA agcensus service provides online access to grid square estimated agricultural census data at various resolutions, for three UK territories, from 1969 to the present day.

- Such a resource forms part of a 'common geographic framework' which allows diverse geo-referenced data and resources to be 'cross-walked' and ultimately utilised.

agcensus:

http://edina.ed.ac.uk/agcensus

Go-Geo!:

http://www.gogeo.ac.uk/

E-MapScholar:

http://edina.ed.ac.uk/projects/mapscholar/index.html

Geo-X-walk:

http://hds.essex.ac.uk/geo-X-walk/