Counting Cows and Cabbages – Web-based Extraction, Delivery and Discovery of Geo-Referenced Data

*by Stuart Macdonald **

About EDINA [1]
EDINA, based at **Edinburgh University Data Library**, is a JISC-funded national data centre. It offers the UK tertiary education and research community networked access to a library of data, information and research resources.

All EDINA services are available free of charge to members of UK tertiary education institutions for academic use, although institutional subscription and end-user registration are required for most services. Services include spatial data services; abstract and indexing bibliographic databases; multimedia and images databases; in addition to a number of geo-related development projects such as geoXwalk, e-MapScholar and Go-Geo!

The spatial data services offered include UKBORDERS (boundary datasets of the United Kingdom), Digimap (Ordnance Survey maps and mapping data) and the EDINA agcensus which allows downloading and the visualisation of grid square agricultural census data from as far back as 1969.

History of the Agricultural Census
Collecting livestock and crop information goes back as far as the Domesday Survey commissioned in December 1085 by William the Conqueror who invaded England in 1066.

In the Middle Ages governments did not collect statistical information for the benefit of the population, or even as a guide to policy but simply for tax and administrative purposes.

In1801 an Agricultural inquiry coincided with the first British Population Census; the reporters were local clergy and a standard form was used to record areas of main crops. This continued into the first half of the nineteenth century where various attempts were made to conduct agricultural inquiries. In general these did not produce a good response. About the same time, the first Statistical Account of Scotland (which was published in twenty-one volumes between 1791 and 1799) was undertaken under the direction of Sir John Sinclair of Ulbster. Based on detailed parish reports, the statistical accounts enumerate and describe such topics as agricultural and industrial production. The second *New Statistical Account* was published between 1834 and 1845. In 1864 Parliament agreed to the collection and publication of agricultural statistics in Great Britain and in 1865 allocated £10,000 to cover the cost. Thus the census began in its modern form in 1866.

The Agricultural Census
The Agricultural Census is conducted annually in June by each of the United Kingdom agriculture departments to help form, monitor and evaluate policy by providing information on the distribution and extent of crop and horticultural production and rearing of livestock.

Each farmer is obliged to declare the agricultural activity on the land via a postal questionnaire. The respective government departments collect the 150 items of data and publish information relating to farm holdings for recognised geographies.

Farm holdings above a certain economic or physical threshold are regarded as Major Holdings with the rest being regarded as Minor. The data provided from the respective government departments and converted by Edinburgh University Data Library correspond to Major holdings only.

Prior to 1998 data for England and Wales was provided by DEFRA. Farmers from both territories completed the same questionnaire. Since 1998 Welsh farmers complete a questionnaire supplied by the Welsh Assembly DEPC. Farmers in Scotland complete a questionnaire as supplied by SEERAD. Thus, due to each government department having responsibility for their respective questionnaire there is not complete comparability over the three territories with regard to census questionnaires. Similarly, due to changes in agricultural policy the content of the questionnaires within the three territories has changed over time. In addition certain census items have been aggregated in recent years to address issues concerning disclosure.

Conversion Algorithms
Areal research in relation to the distribution of agricultural census data was carried out at the University of Edinburgh in the 1970's and 1980's by Professor Terry Coppock and

Jack Hotson, with the co-operation with MAFF and ADAS. The level of publication of the Agricultural Census data was the parish summary. Coppock and Hotson developed algorithms to redistribute the parish summary data into 5km and 10km grid square estimates, taking into account potential land uses. This was done for the following reasons

- the geographies (e.g. Scottish parishes, Welsh communities) vary in size, shape, the land use capability
- parish summaries may under- or over-report agriculture activity e.g. a farmer need make only one return, even if some land or livestock are remote from the main holding
- the census returns are a 'snapshot' of activity on 1 June
- grid square data aligned to the National Grid facilitate analysis of data over a number of years and with other data sets.
- the format of data obtained from the government departments is potentially disclosive

The key to transforming the Agricultural Census data into grid square data was the definition of each geography (parish) as 1km squares. This Framework was used in conjunction with a 7-fold classification of the land-use of the same 1km grid squares called the Land-use Framework. The resulting distribution of the data gave a good estimate at 5km level of "what was likely to be where", as well as protecting farmers' confidentiality.

Migration to a Web service

Up until recently data had to be extracted from a mainframe or local server, using a set of command driven extraction programs. The data could then be mapped using the interactive Gridmap utility, written in-house by Alison Bayley, building on the raster-type CAMAP software written by Jack Hotson for data retrieval and grid mapping.

However the emergence of desktop GIS and web technologies has enabled the service to develop further, making it potentially available to a newer and wider audience.

New EDINA agcensus

In spring 2004 preliminary work began on processing and reformatting existing data into a grid square format suitable for importation into and delivery from a MySQL database. Data from census years 1969, 1976, 1981, 1988 and 1994 was reworked from existing grid formatted data, while year 2000 data was converted directly into a suitable format from the area-specific data provided by the respective government departments. This allows analysis of change over time with intervening and more recent census data to be processed in due course.

Existing data processing algorithms were re-written in Java to allow for easy maintenance and migration between machines. The interface uses Java Servlets and JSP technology to enable clear presentation and access to the data. The options presented in the interface reflect the variation in censuses from country to country and from year to year.

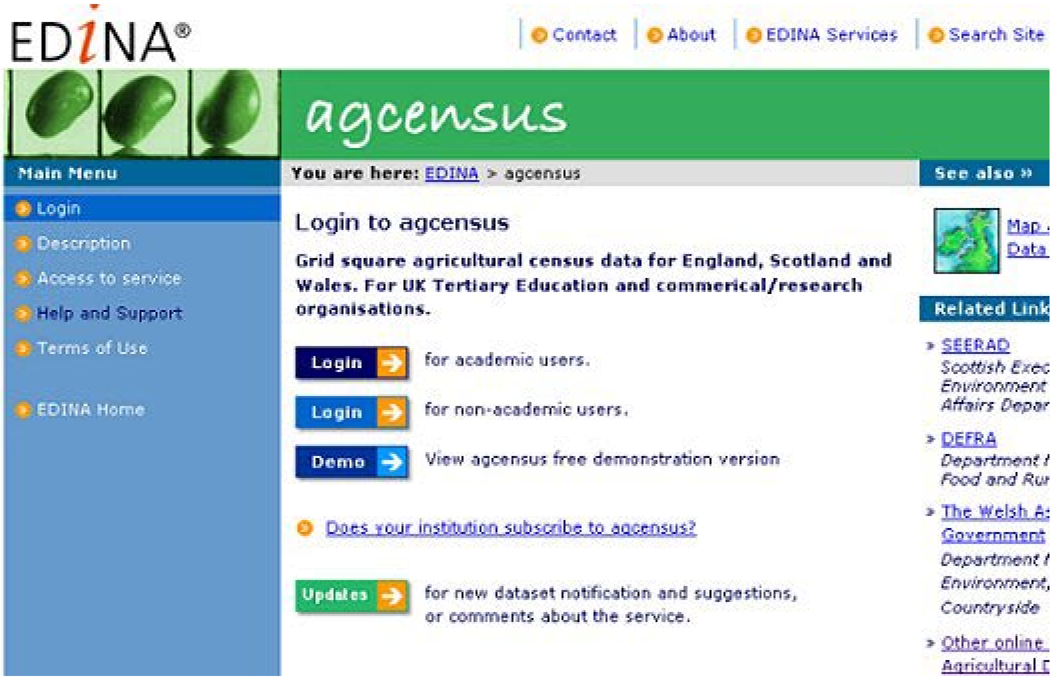
The post-processing database holds a table for each country, year and grid resolution combination, for which metadata is held in two lookup tables. The lookup tables hold information about what tables are available, which census items occur in each country/year combination and text descriptions, groupings and units of analysis.

The data visualisation component of the new service uses image generation servlets which present the data in map form for the area, census item, year, grid size and bandings specified.

Decisions were made (and problems resolved) regarding the provision of an alternative visualisation for context mapping, and the use and development of the area select tool; allowing for both javascript and non-javascript functionality in addition to accessibility. It was also decided that the service would be more responsive if data were held in aggregate form rather than aggregated for each data request, after weighing up the issues of storage versus ‘on the fly’ delivery.

Two Bartholomew’s raster datasets (1:200,000 and 1:800,000) were used as the context mapping for Great Britain. Land use data from the 1980’s (for Scotland, England and Wales) were used to provide an alternative context.

The new EDINA agcensus service was launched on October 1st 2004 offering grid square Agricultural Census data to 3 client communities: the academic community (via an annual Athens authenticated subscription); commercial organisations, and research/policy makers (both via an EDINA controlled authorisation and authentication on a ‘per project’ basis) (see Fig 1).



Login to agcensus

Grid square agricultural census data for England, Scotland and Wales. For UK Tertiary Education and commerical/research organisations.

Login for academic users.

Login for non-academic users.

Demo View agcensus free demonstration version

Does your institution subscribe to agcensus?

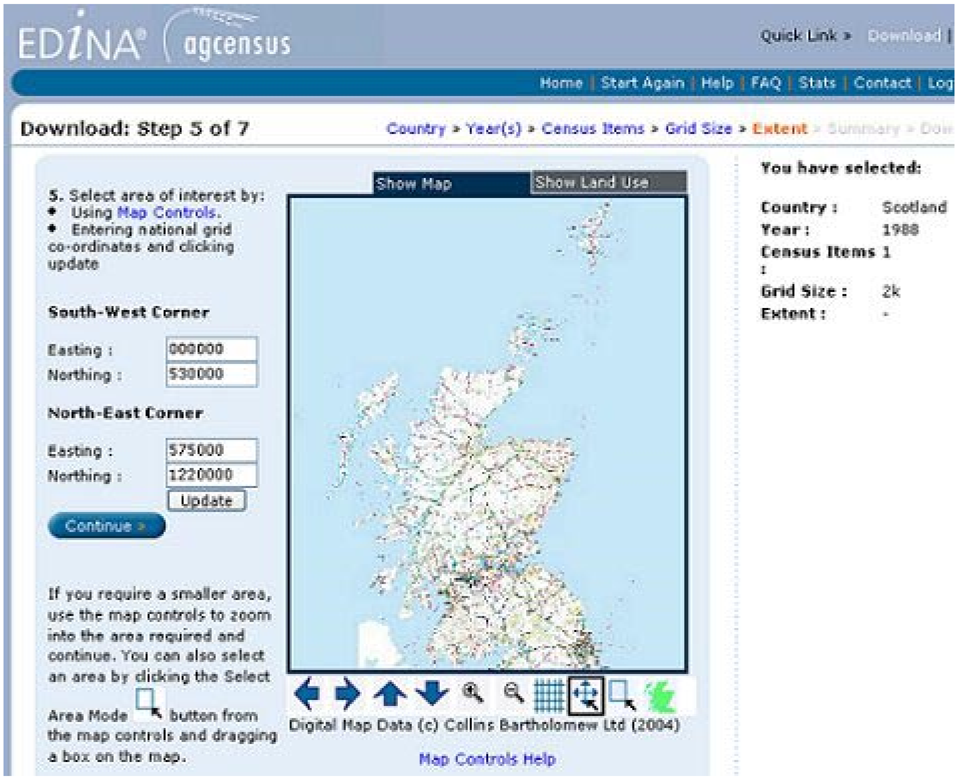
Updates for new dataset notification and suggestions, or comments about the service.

Fig 1: The EDINA agcensus interface offering access to academic and non-academic customers in addition to a free demonstration version of the service

A free visualisation demo service containing all census items at 10km resolution for the most recent census year held enables potential users to preview distribution maps of chosen census items although no data download is offered.

Authorised users are presented with a simple interface offering two routes into the data, namely **data download** (ASCII delimited comma separated values) and **data visualisation** (distribution maps of census items) via a 7 step access procedure:

- Select country from Scotland, England, Wales (also GB for small subset of items)
- Select year from1969, 1976, 1981, 1988, 1994, 2000
- Select census item(s) – at present only one census item can be visualised at a time. All (or a subset there of) of the census items for a chosen year at a chosen resolution can be downloaded
- Select grid size (level of aggregation) from 2km, 5km and 10km
- Select extent of data coverage (see Fig. 2)


EDINA agcensus
Quick Link » Download |
Home | Start Again | Help | FAQ | Stats | Contact | Log
Country » Year(s) » Census Items » Grid Size » Extent » Summary » Dow
Show Map
Show Land Use
Update
Continue »
Digital Map Data (c) Collins Bartholomew Ltd (2004)
Map Controls Help

Download: Step 5 of 7

5. Select area of interest by:
- Using Map Controls.
- Entering national grid co-ordinates and clicking update

South-West Corner

Easting :	000000
Northing :	530000

North-East Corner

Easting :	575000
Northing :	1220000

If you require a smaller area, use the map controls to zoom into the area required and continue. You can also select an area by clicking the Select Area Mode button from the map controls and dragging a box on the map.

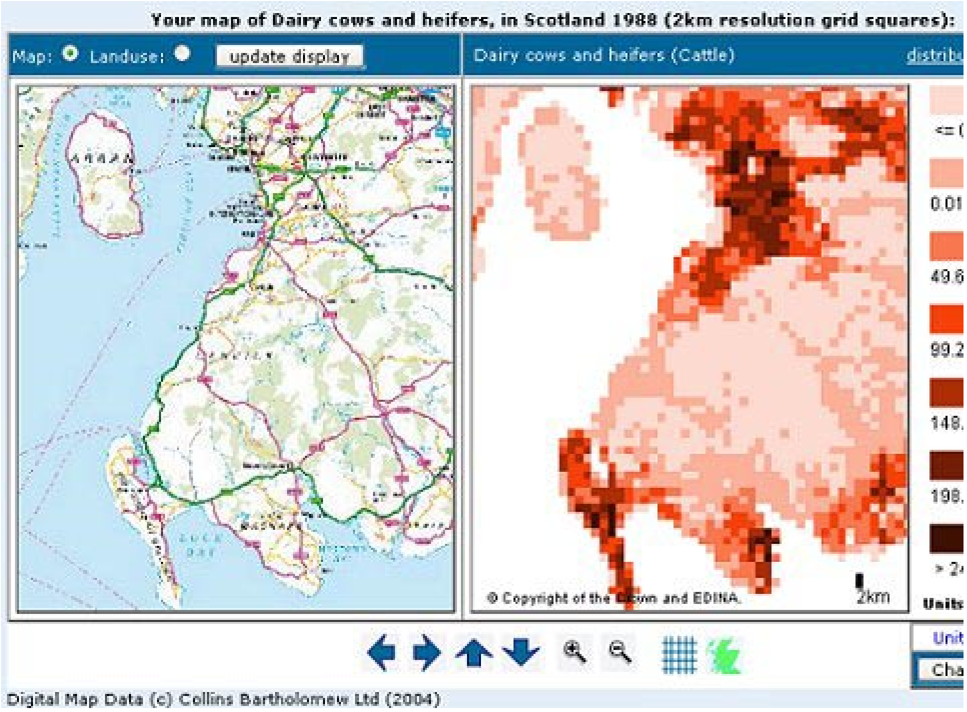
You have selected:

Country :	Scotland
Year :	1988
Census Items :	1
Grid Size :	2k
Extent :	-

Fig 2: For small area analysis use the extent tool or enter British National Grid co-ordinates. Toggle between context map and land use map.

- Data selection summary (allows user to change chosen parameters)
- Download or visualise selected census data.

As an example, cattle distribution for the south west area of Scotland has been visualised (see Fig. 3). The image itself can be downloaded as a GIF file by right clicking on the image or printed off as a hard copy. In addition the unit bandings can be customised and the land use data for the specified area can be viewed. After visualising the chosen item for the selected area the corresponding data can be downloaded by clicking on the appropriate icon.


Your map of Dairy cows and heifers, in Scotland 1988 (2km resolution grid squares):

© Copyright of the Crown and EDINA.

Digital Map Data (c) Collins Bartholomew Ltd (2004)

Fig 3: Distribution map of dairy cows and heifers in Scotland, 1988 at 2km resolution.

Service Issues

Prior to service launch the following key points were addressed:

Authentication. This was achieved via the Athens Access Management system which provides users from the UK tertiary education community with single sign-on to numerous web-based services. Non-academic authorisation and authentication to the service is managed by the EDINA helpdesk.

Documentation. Availability of online user guides, questionnaires and publicity material (and hardcopy on request).

Field Trialing. Field trials of the service were conducted by the Scottish Agricultural College using Nielsen-Norman Usability tests to precipitate feedback. Feedback from in-house testing was also implemented into the interface and functionality of the service.

Subscription Model. Unlike other EDINA services *agcensus* is an exception with regard to subscription in that commercial/ policy/research organisations can gain access to the data. The JISC banding structure was employed for the academic audience. However those from the aforementioned non-academic institutions subscribe on a per project basis (allowing institutional access). Individuals from both academic and non-academic organisations unable to raise relevant subscription costs also have the option to pay on a per project basis.

Training/Outreach. Structured training events and workshops will be organised to both publicise and demonstrate the service. This will be done in conjunction with the EDINA Training Officer through liaison with external institutions. Online training materials will be made available via the EDINA agcensus website.

User Support. The EDINA agcensus service is supported by the EDINA helpdesk which adheres to a set of service level definitions. Enquiries are dealt with via telephone and email with service downtime, alerts, upgrades etc being posted on the EDINA agcensus website. Technical and in-depth service support are also available.

Accessibility. An Accessibility Statement explains EDINA's policy of working towards maximum accessibility for all users to all services.

Further Developments

There are a number of developments being investigated for Version 2 of the service. These include:

- introducing data that complement the grid square agricultural census datasets such as gridded meteoro logical data, species data, historic census data (at present data for the 1871 Scottish agricultural parishes are being digitised for processing)
- online visualisation of change over time for a chosen census item
- statistical reporting to provide summary statistics and enable rudimentary numerical analysis of the census data and predictive modelling
- combining more than one census item together for visualisation or download
- a teaching dataset for use in learning and teaching in the classroom, laboratory etc
- other visualisations such as histograms, pie charts, bar charts

The Bigger Picture

As a UK national data centre, EDINA engages in both projects and services, the former being geared to development activities which inform and develop the operation of EDINA national services, either producing new services or improvement in existing services. Projects are generally externally-funded and often in partnership with other institutions.

Three projects funded under the JISC 5/99 Programme relate to web delivery of spatial data to the UK HE/FE sector, and build on the national significance of Digimap (which delivers access to Ordnance Survey mapping) and UKBORDERS (digitised boundaries). These are Go-Geo!, geoXwalk and E-MapScholar.

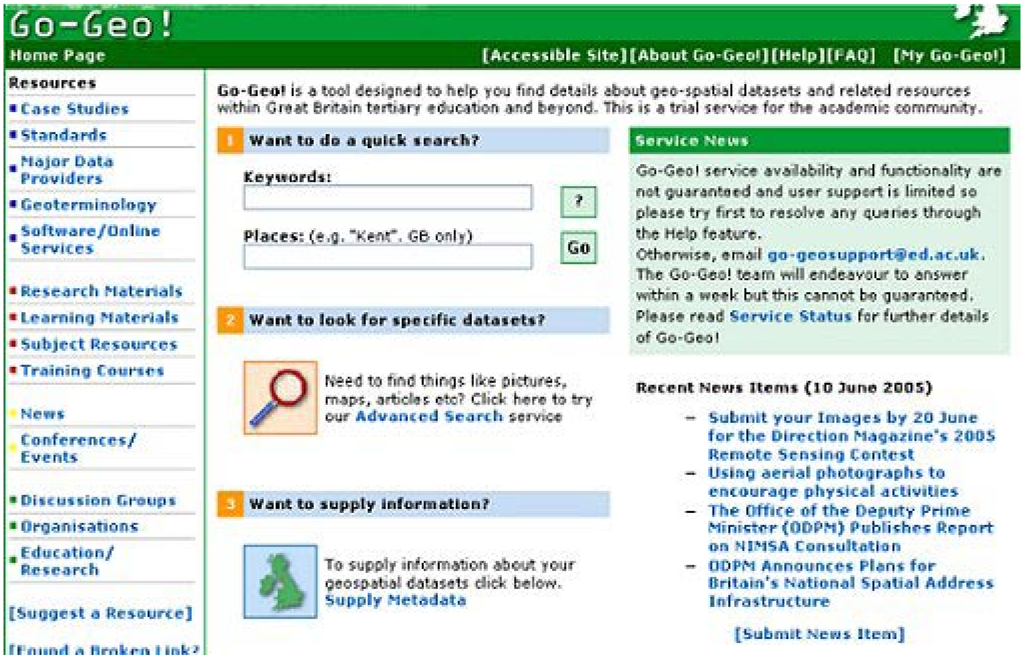
Go-Geo!

With increasing amounts of spatial data being created within Higher Education, demand for managed access to this data is growing with GIS tools becoming more commonly available. However, two barriers confront the potential user of spatial data:

- how to find out what datasets exist
- how to ascertain their quality and suitability for use.

These barriers can be overcome by comprehensive, standardised metadata, available through the web-searchable portal.

Go-Geo! is a JISC-funded project run jointly by the EDINA and the UK Data Archive. As a Z39.50 compliant resource discovery tool it allows identification and retrieval of metadata describing the content, quality, condition and other characteristics of spatial data within and beyond the UK HE community. The Metadata Profile originally employed by the Go-Geo! project was based on UK NGDF Guidelines and the ISO 19115 Geographic Information Metadata Standard which was adopted in March 2003 and mapped recently to Dublin Core. Go-Geo! also acts as the academic node of the UK **GIgateway** service (hosted at EDINA) and can go beyond discovery to provide direct access to data in some cases (see Fig. 4).



Go-Geo!

Home Page [Accessible Site][About Go-Geo!][Help][FAQ] [My Go-Geo!]

Resources
- Case Studies
- Standards
- Major Data Providers
- Geoterminology
- Software/Online Services
- Research Materials
- Learning Materials
- Subject Resources
- Training Courses
- News
- Conferences/ Events
- Discussion Groups
- Organisations
- Education/ Research

[Suggest a Resource]

[Found a Broken Link?

Go-Geo! is a tool designed to help you find details about geo-spatial datasets and related resources within Great Britain tertiary education and beyond. This is a trial service for the academic community.

1 **Want to do a quick search?**

Keywords: ?

Places: (e.g. "Kent". GB only) Go

2 **Want to look for specific datasets?**

Need to find things like pictures, maps, articles etc? Click here to try our **Advanced Search** service

3 **Want to supply information?**

To supply information about your geospatial datasets click below.
Supply Metadata

Service News

Go-Geo! service availability and functionality are not guaranteed and user support is limited so please try first to resolve any queries through the Help feature.
Otherwise, email go-geosupport@ed.ac.uk.
The Go-Geo! team will endeavour to answer within a week but this cannot be guaranteed.
Please read **Service Status** for further details of Go-Geo!

Recent News Items (10 June 2005)

- Submit your Images by 20 June for the Direction Magazine's 2005 Remote Sensing Contest
- Using aerial photographs to encourage physical activities
- The Office of the Deputy Prime Minister (ODPM) Publishes Report on NIMSA Consultation
- ODPM Announces Plans for Britain's National Spatial Address Infrastructure

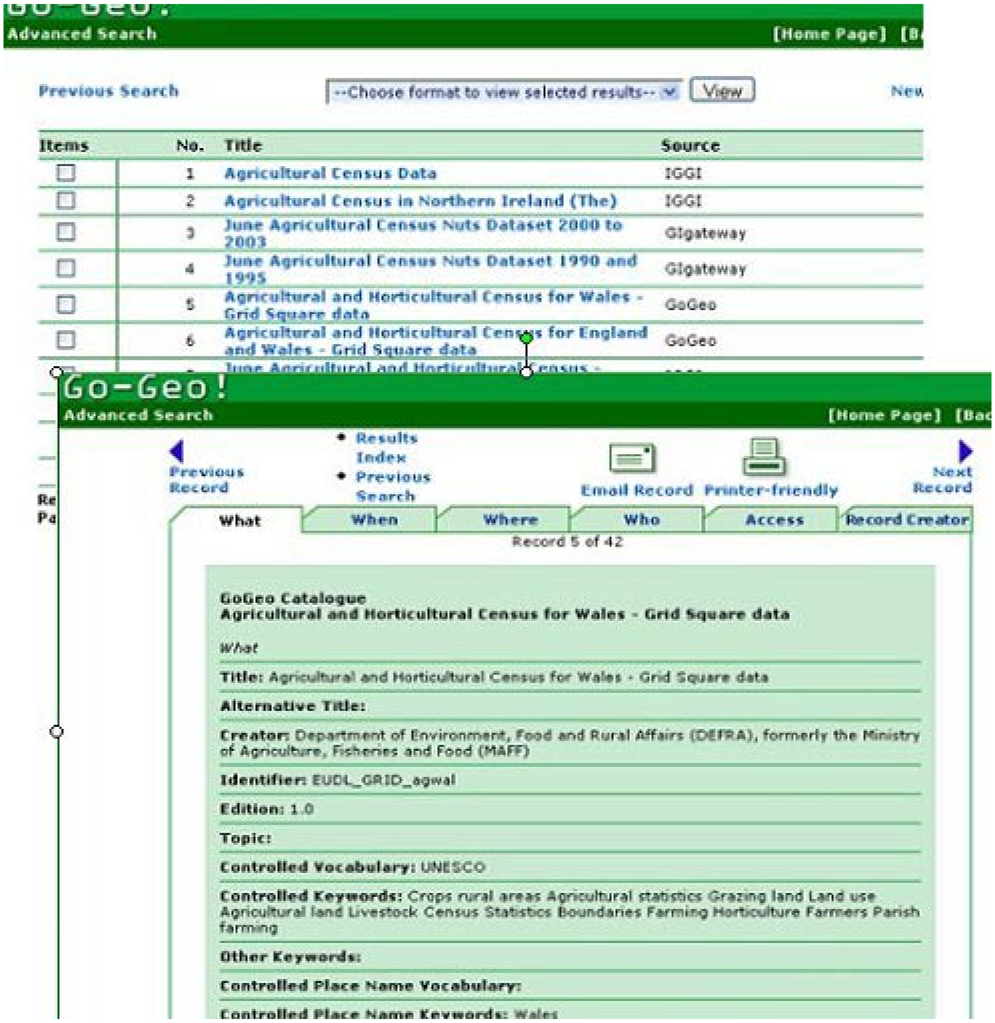
[Submit News Item]

Fig 4: Go-Geo! portal offers both simple and advanced search options in addition to a 'library' of geo-resources including news items, case studies, learning materials, data providers, training courses and discussion groups.

A simple keyword search for e.g. *agricultural census* retrieves a number of results allowing the metadata for chosen records to be viewed (see Fig. 5).

With the advanced search facility searches can be restricted by data type (maps, images, datasets, reference material, projects etc), location, text, data range.

In addition to cross-searching spatial databases from major data producers Go-Geo! also extends the data discovery function by providing access to a 'library' of other related resources of use to the user. These resources can be either local to the portal or found by searching the JISC Information Environment and other online information services.



Items	No.	Title	Source
□	1	Agricultural Census Data	IGGI
□	2	Agricultural Census in Northern Ireland (The)	IGGI
□	3	June Agricultural Census Nuts Dataset 2000 to 2003	GIgateway
□	4	June Agricultural Census Nuts Dataset 1990 and 1995	GIgateway
□	5	Agricultural and Horticultural Census for Wales - Grid Square data	GoGeo
□	6	Agricultural and Horticultural Census for England and Wales - Grid Square data	GoGeo
		June Agricultural and Horticultural Census -	

Record 5 of 42

GoGeo Catalogue
Agricultural and Horticultural Census for Wales - Grid Square data

What

Title: Agricultural and Horticultural Census for Wales - Grid Square data

Alternative Title:

Creator: Department of Environment, Food and Rural Affairs (DEFRA), formerly the Ministry of Agriculture, Fisheries and Food (MAFF)

Identifier: EUDL_GRID_agwal

Edition: 1.0

Topic:

Controlled Vocabulary: UNESCO

Controlled Keywords: Crops rural areas Agricultural statistics Grazing land Land use Agricultural land Livestock Census Statistics Boundaries Farming Horticulture Farmers Parish farming

Other Keywords:

Controlled Place Name Vocabulary:

Controlled Place Name Keywords: Wales

Fig 5: Retrieved results and individual record complete with 'what, when, where' metadata tabs

geoXwalk
The main purpose of this project was to provide a shared service within the JISC Information Environment (IE) that can support geographic searching. At present each information provider or service adopts different geographic coding principles (such as postcode, place name, grid reference). Thus the creation of an online Z39.50 compliant British and Irish gazetteer would facilitate a unified entry point into geographical searching in addition to providing researchers and teachers with an online reference tool.

The geoXwalk gazetteer itself contains a list of place names with their associated spatial location expressed in several ways (e.g. latitude and longitude co-ordinates etc). It also classifies features into types such as cities as areas, rivers as lines etc and stores an appropriate spatial 'footprint' against each feature. This introduces transparency to a geographic search by allowing the 'cross-walking' of these different geographies in addition to allowing searches to be conducted on a proximity/distance basis.

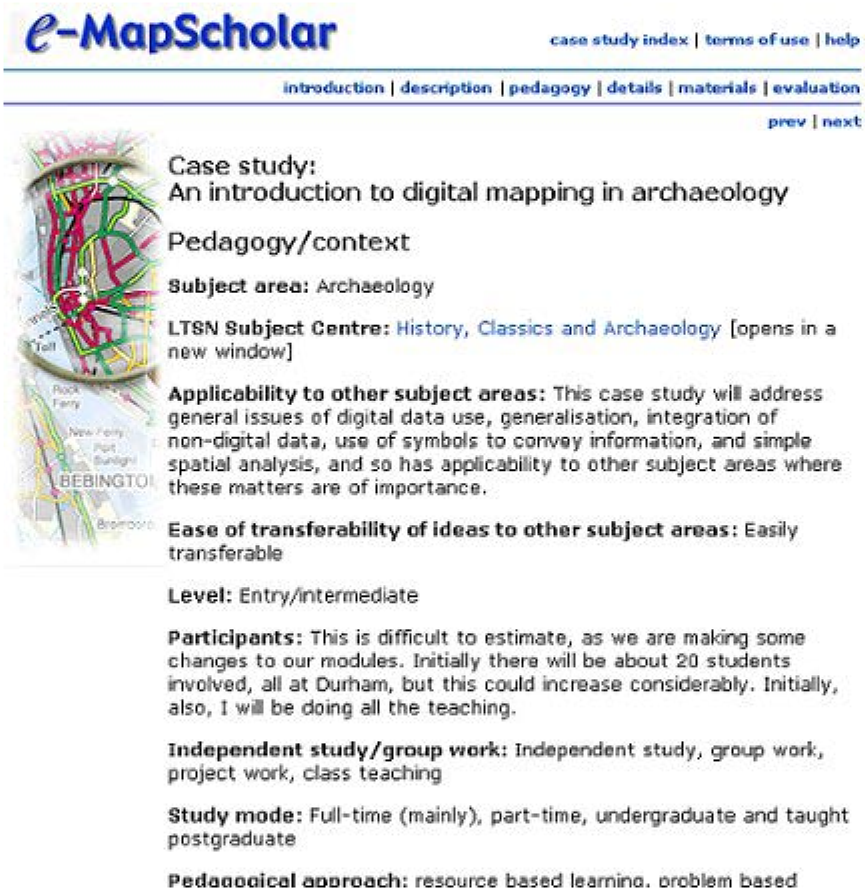
Integral to such a project was the need for a **geoparser**, software than can 'read' and automatically identify place names in an electronic document (e.g. a resource description or digitised historical document). Such identified place names can then be compared against the geoXwalk gazetteer entries thus providing access to 'alternate' geographies by the assignment of geo-tags (e.g. a grid reference) to implicit geo-referenced material (e.g. a place name). Thus the combination of the parser and the digital gazetteer has potential for powerful geographic based searching across a range of otherwise disparate resources such as those contained within the JISC IE.

E-MapScholar
The third of EDINA's JISC-funded spatial projects, e-MapScholar, is a learning and teaching project.

From an earlier project (the JISC Electronic Libraries (eLib) Digimap Project) it was identified that there was a skills/ concepts gap between creating a map and downloading and using digital map data in a GIS. Thus the aim of e-MapScholar was to fill this gap by developing tools to promote the use of spatial data, including OS digital map data available from the EDINA Digimap service, initially within tertiary education. However the model could be applied to other levels of education. It would support both those learners who need to progress to using a GIS in addition to those whose needs are more straightforward.

This spatial data literacy project has four components:

Teaching case studies which consist of the data and materials used by the learners, along with descriptions of how the data and learning materials have been integrated into a variety of disciplines, and evaluations by staff and students (see Fig. 6).


e-MapScholar
case study index | terms of use | help
introduction | description | pedagogy | details | materials | evaluation
prev | next
Toll
BEBINGTO

Case study:
An introduction to digital mapping in archaeology

Pedagogy/context

Subject area: Archaeology

LTSN Subject Centre: History, Classics and Archaeology [opens in a new window]

Applicability to other subject areas: This case study will address general issues of digital data use, generalisation, integration of non-digital data, use of symbols to convey information, and simple spatial analysis, and so has applicability to other subject areas where these matters are of importance.

Ease of transferability of ideas to other subject areas: Easily transferable

Level: Entry/intermediate

Participants: This is difficult to estimate, as we are making some changes to our modules. Initially there will be about 20 students involved, all at Durham, but this could increase considerably. Initially, also, I will be doing all the teaching.

Independent study/group work: Independent study, group work, project work, class teaching

Study mode: Full-time (mainly), part-time, undergraduate and taught postgraduate

Pedagogical approach: resource based learning, problem based

Fig. 6: An example of a teaching case study "An introduction to digital mapping in archaeology"

Online learning and teaching materials such as tutorials have been developed with interactive tools which enable users to develop skills in the use of digital map data and knowledge of spatial data concepts such as integration and visualisation.

A **teaching content management system** has been developed which allows teaching staff to customise and re-purpose the online learning materials to suit their curriculum.

A **virtual work placement** has been designed in which students can carry out an assessment of the visual impact of wind turbines at the Nant Carfan development in Wales. This provides the opportunity for learners to develop workplace-related skills in the use of spatial data, using problem-based learning techniques.

At present the JISC are funding a follow-on project to look at ways in which they might be able to make the products from the e-MapScholar available in a service environment.

Summary

The EDINA agcensus service forms part of a growing number of geo-data resources utilised within UK academia and has evolved using web technologies to enable data access to an expansive audience. Projects related to spatial data services such as Go-Geo!, geoXwalk and eMapScholar aim to raise awareness of geo-data and associated resources. Additionally they enhance access to, and use of geo-resources to those both within and beyond the academic sector. Such resources highlight the role of data service providers, such as EDINA, in offering and strengthening networked access to a collection of data, information and digital materials to the UK tertiary education and research community.

Comments on the EDINA agcensus service are welcome. Email the author at: stuart.macdonald@ed.ac.uk.

Footnotes

[1] All URLs and acronyms are listed here in the Appendix.

Appendix:

URLs

EDINA National Data Centre:	http://edina.ac.uk
EDINA Digimap:	http://edina.ed.ac.uk/digimap
EDINA UKBORDERS:	http://edina.ed.ac.uk/ukborders
EDINA agcensus:	http://edina.ac.uk/agcensus
The Domesday Book Online:	http://www.domesdaybook.co.uk
Statistical Accounts of Scotland:	http://edina.ac.uk/stat-acc-scot
Go-Geo!	http://www.gogeo.ac.uk
geoXwalk	http://www.geoxwalk.ac.uk
e-MapScholar	http://edina.ac.uk/projects/mapscholar/index.html
GIgateway	http://www.gigateway.org.uk

Acronyms

JISC	Joint Information Systems Committee
DEFRA	Department of the Environment, Forestry and Rural Affairs (previously MAFF)
SEERAD	Scottish Executive Environment and Rural Affairs Department
DEPC	Department for Environment, Planning and Countryside
MAFF	Ministry for Agriculture, Fisheries and Food
ADAS	Agricultural Development and Advisory Service
GIS	Geographic Information Systems
NGDF	National Geospatial Data Framework
OS	Ordnance Survey

* Paper presented at the IASSIST Conference, Madison, May 2004, by Stuart Macdonald. Contact: Stuart Macdonald, EDINA National Data Centre & Edinburgh University Data Library, Tel. 0131-650-3304 | e-mail: Stuart.Macdonald@ed.ac.uk