

Introduction to GEMDE: 'Grid Enabled ethnic Minority Data Resources'

Paul Lambert, Tom Doherty, Susan McCafferty, and others, 28th January 2010, Univ. Stirling

Presented to DAMES workshop on 'Data on ethnicity in social survey research'

1

GESDE: Grid Enabled Specialist Data Environments

- Facilities for collecting together, and distributing, specialist data resources
 - Occupations: GEODE project began 2005
 - Education and Ethnicity: GEEDE and GEMDE began Feb. 2008
- Capacity building aims: improving use of measures of these concepts by
 - *improving access to relevant information*
 - *providing training / advice on good practice*

2

GEODE: Grid Enabled Occupational Data Environment

Current facilities

- Deposit data in any format with abstract / basic description
 - Upload dataset and/or supply uri
- Search and browse facilities for deposited data
 - (a little difficult to use)
- 'Occupational matching' routine on plain text files
- Specialist curation in xml (DDI) required to fully integrate with file matching routines

GEODE: Organising and distributing specialist data resources (on occupations)

The screenshot shows a web browser window titled "GridSphere Portal - Windows Internet Explorer provided by University of Stirling". The address bar shows the URL: <http://geode.cs.stir.ac.uk:8080/gridsphere/gridsphere?cid=Deposit+Data&JavaScript=enabled>. The page content includes a navigation menu with items like "Welcome", "Administration", "Feedback", "Deposit data", "GEODE Search", "GEODE Browse", and "Link data resources". The main content area is titled "Grid Enabled Occupational Data Environment" and displays a form for "Meta Data Information". The form fields are as follows:

Meta Data Information	
Title	CAMSIS scale for Ireland from 1996
Supplier's Information	
Name	Paul Lambert
Affiliation	University of Stirling
Email	paul.lambert@stirling.a
Original Creator's Information	
Name	Paul Lambert; Richard Layte
Original Production Date	2002
URL of your resource	http://www.camsis.stir.ac.uk/Data/Ireland96.html
Abstract	Zip file containing the SPSS and plain text data files for the CAMSIS scales for Ireland using 1996 data, coded to the Irish national occupational classification (CSO-96) and the International Standard Classification of Occupations 1988 (ISCO88).
Occupational Classification Used	CSO-96 Chief Scientists Office 1996 Classification scheme ISCO-89 International Standard Classification of Occupations 1988

GEODE: Grid Enabled Occupational Data Environment

Plans / requirements

- Improved searching / browsing of deposited data
- Increased numbers of users / depositors
- Automated / alternative data curation arrangements in DDI3
- File linking with SPSS / Stata

- More open access online resources
 - Listings of OUGs etc
 - Experts Wiki / FAQs
 - Stata format data files online, e.g
http://www.camsis.stir.ac.uk/downloads/CAMSIS_downloads.html

Ethnicity and the DAMES project

- *Hard subject to collate information on*
 - Few recognisable 'ethnic unit groups'
 - Limited previous 'data management' reflection
 - Very few published databases on ethnicity
 - Important question of sparse distributions
 - Dynamic, & rapidly expanding

- Likely role is to give new guidance on emerging strategies for analysing and exploiting data

GESDE(i): Basic access to data

Services to..

- search for and identify suitable information resources
{Liferay portal and iRODS file connection}
- allow merging these resources with own data
{Non-trivial consideration – complex micro-data subject to security constraints}
- Constructing new standardized resources for UK and major cross-national surveys
E.g. Effect proportional scales for ethnic groups and educational qualifications across countries and over time
CAMSIS scales for educational homophily (cf. www.camsis.stir.ac.uk)

7

GESDE(ii): Depositing data

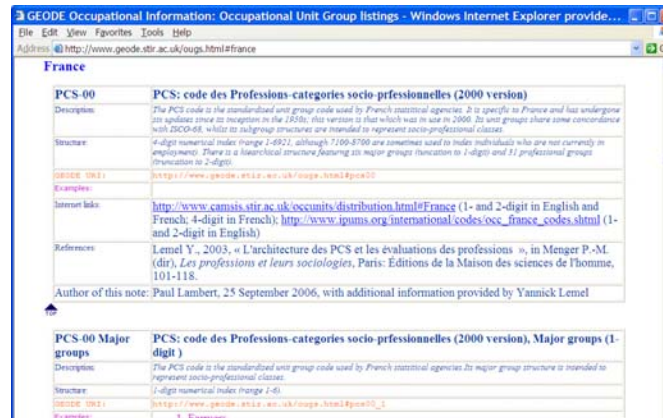
Services to...

- Allow researchers to deposit specialist information resources to be immediately visible to others
- Collect basic metadata via proforma, option of adding extended metadata (DDI structure)
{Motivations are altruism; citations; reduced burdens}
{Quality control through site rankings, expert inputs}

8

The GEODE model for GEMDE?

- Occupational Information Resources
- Occupational Unit Groups



Occupational information resources: small electronic files about OUGs...

	Index units	# distinct files (average size kb)	Updates?
CAMSIS, www.camsis.stir.ac.uk	Local OUG*(e.s.)	200 (100)	y
CAMSIS value labels www.camsis.stir.ac.uk	Local OUG	50 (50)	n
ISEI tools, home.fsw.vu.nl/~ganzeboom	Int. OUG	20 (50)	y
E-Sec matrices www.iser.essex.ac.uk/esecc	Int. OUG*(e.s.)	20 (200)	n
Hakim gender seg codes (Hakim 1998)	Local OUG	2 (paper)	n

E.g.: UK 1980 CAMSIS scales and CAMCON classes (www.camsis.stir.ac.uk)

*gb91co80.sav [DataSet1] - SPSS Data Editor

opcod	stdempst	ukempst	mcamsis	fcamsis	co80	co80str	seg	rgsc	casmin	var
977	0	0	41.7	40.0
978	-6	1	41.7	39.6	53.01	053.1	.70	.0	.0	.
979	-5	2	41.7	39.6	53.01	053.1	.70	.0	.0	.
980	3	3	41.7	39.6	53.01	053.1	.70	.0	.0	.
981	-4	4	41.7	39.6	53.01	053.1	.70	.0	.0	.
982	-3	5	41.7	39.6	53.01	053.1	.70	.0	.0	.
983	-2	6	39.1	52.6	52.01	052.1	8.00	4.0	8.0	.
984	-1	7	41.8	39.6	53.01	053.1	10.00	5.0	10.0	.
985	0	0	45.2	38.6
986	-6	1	46.8	37.9	53.02	053.2	.30	.0	.0	.
987	-5	2	46.8	37.9	53.02	053.2	.30	.0	.0	.
988	3	3	55.6	66.6	53.02	053.2	12.00	6.0	6.0	.
989	-4	4	46.8	37.9	52.02	052.2	.60	.0	.0	.
990	-3	5	46.8	37.9	52.02	052.2	.60	.0	.0	.
991	-2	6	39.1	37.9	52.02	052.2	8.00	4.0	8.0	.
992	-1	7	45.2	37.9	53.02	053.2	11.00	6.0	10.0	.
993	0	0	54.7	44.5
994	-6	1	54.7	44.5	55.01	055.1	.70	.0	.0	.
995	-5	2	54.7	44.5	55.01	055.1	.70	.0	.0	.

Our approach to GEMDE



-A service for MUGs and MIRs...
 - Define/register 'Minority Unit Groups'
 - Define/register 'Minority Information Resources'
 - Explore data resources and obtain help in approaching analysis of complex, sparse data

DAMES Project Web Page - Windows Internet Explorer provided by U...

Address <http://www.dames.org.uk/gemde/>

FREQUENTLY ASKED QUESTIONS

- 1) What's a MUG?
 - We use this acronym for '**Minority Unit Group**'. This is our own terminology which draws parallels with the better known 'Occupational Unit Group' (see [listing of OUGs](#) at the GEODE site). A Minority Unit Group is any **systematic listing of the categories of a measure of ethnicity**.
 - Example: In the UK, the Census ethnic group questions are well known examples of MUGs. E.g., the 2001 Census ethnic group question for England and Wales constitutes a MUG with the following principal categories: "British", "Irish", "Other White", "White and Black Caribbean", "White and Black African", "White and Asian", "Other Mixed", "Indian", "Pakistani", "Bangladeshi", "Other Asian", "Black Caribbean", "Black African", "Other Black", "Chinese", "Other Ethnic Group" (see e.g. [the SARS listing of this variable](#)).

What's a MIR?

- We use this acronym for '**Minority Information Resource**'. This is our own terminology. By a MIR, we mean any piece of

Local intranet

13

Microsoft Excel

ethhp_s_census2001.xls [Compatibility Mode]

	B	C	D	E	F	G	H	I	J
3	2001 Census:	Ethnic group for Scotland				Source:	http://www.ccsr.ac.uk/sars/		
4							(accessed 26/1/2010)		
5		1	White Scottish						
6		2	Other white British						
7		3	White Irish						
8		4	Other white						
9		5	Any mixed background						
10		6	Indian						
11		7	Pakistani						
12		8	Bangladeshi						
13		9	Other South Asian						
14		10	Caribbean						
15		11	African						
16		12	Black Scottish or other black						
17		13	Chinese						
18		14	Other ethnic group						

What's a MIR?

- '**Minority Information Resource**'
 - This is our own terminology. By a MIR, we mean any piece of information which supplies systematic data on a minority unit group (MUG) classification. We've used this term to be deliberately similar to the phrase 'Occupational Information Resources' that we used on GEODE
- E.g. summary statistical data about the categories from and documentation or information
- E.g. recodings which have been used in a particular study
 - Social scientists are not in general aware of the existence of MIRs (cf. wide use of popular Occupational Information Resources). In GEMDE we seek to publicise little known resources and promote their uptake: We argue that better communication and dissemination of MIRs is in fact an important step towards better scientific practice of replication and standardisation of research.
- ***In our terms, every MIR necessarily links to a MUG (but not every MUG has a MIR).***

The GEMDE prototype

'Liferay portal' with access to MUGs and MIRs

- **Current facilities**

- Shibboleth access
- Deposit MUGs/MIRs
- Search/browse deposited resources
- Feedback on resources (user ratings)

- **Still to come**

- Additional guest access
- Review live data (e.g. pooled LFS records)
- Expert and user quality ratings

=> ...Over to Tom...

15