Handling Occupational Information and Introduction to GEODE

GEODE – www.geode.stir.ac.uk

Grid Enabled Occupational Data Environment
Session 1 of GEODE Project workshop, 16th January 2007

Paul Lambert, Larry Tan, Ken Turner, & Vernon Gayle	**University of Stirling**
Ken Prandy	**Cardiff University**
Richard Sinnott	**University of Glasgow**

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

GEODE, 16 Jan 2007

NCeSS

Grid Enabled Occupational Data Environment

1. Handling Occupational Information
   ➤ some principles and problems

GEODE activities and illustrations:

2. Occupational Information Depository
3. Access to occupational information

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

GEODE, 16 Jan 2007

NCeSS

Why occupational analyses?

"A man's work is as good a clue as any to the course of his life and to his social being and identity" (Hughes, 1958)

"Nothing stamps a man as much as his occupation. Daily work determines the mode of life.. It constrains our ideas, feelings and tastes" (Goblot, 1961)

"The backbone of the class structure, and indeed of the entire reward system of modern Western society, is the occupational order" (Parkin, 1972)

(Quotes as reproduced in Coxon and Jones 1978; Crompton 1998)

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

GEODE, 16 Jan 2007

NCeSS

Context

- **Occupational information crucial to social science investigation**
  - Social class and social classifications
  - Employment statistics
  - Occupations and economics

- **Most nations have facilities for collecting micro-data with occupational codes:**
  - www2.warwick.ac.uk/fac/soc/ier/publications/software/cascot/

- **We lack *accessible and standardised* facilities for dealing with occupational micro-data**

GEODE, 16 Jan 2007

Occupational information resources: small electronic files…

	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/esec	Int. OUG*(e.s.)	20 *(200)*	n
Hakim gender seg codes (Hakim 1998)	Local OUG	2 *(paper)*	n

GEODE, 16 Jan 2007

For example: ISCO-88 Skill levels classification

and: UK 1980 CAMSIS scales and CAMCOM classes

	opcod	stdempst	ukempst	mcamsis	fcamsis	co80	co80str	seg	rgsc	casmin	var
977	123	0	0	41.7	40.0						
978	123	-6	1	41.7	39.6	53.01	053.1	.70	.0	.0	
979	123	-5	2	41.7	39.6	53.01	053.1	.70	.0	.0	
980	123	3	3	41.7	39.6	53.01	053.1	.70	.0	.0	
981	123	-4	4	41.7	39.6	53.01	053.1	.70	.0	.0	
982	123	-3	5	41.7	39.6	53.01	053.1	.70	.0	.0	
983	123	-2	6	39.1	52.6	52.01	052.1	8.00	4.0	8.0	
984	123	-1	7	41.8	39.6	53.01	053.1	10.00	5.0	10.0	
985	124	0	0	45.2	38.6						
986	124	-6	1	46.8	37.9	53.02	053.2	.30	.0	.0	
987	124	-5	2	46.8	37.9	53.02	053.2	.30	.0	.0	
988	124	3	3	55.6	66.6	53.02	053.2	12.00	6.0	6.0	
989	124	-4	4	46.8	37.9	52.02	052.2	.60	.0	.0	
990	124	-3	5	46.8	37.9	52.02	052.2	.60	.0	.0	
991	124	-2	6	39.1	37.9	52.02	052.2	8.00	4.0	8.0	
992	124	-1	7	45.2	37.9	53.02	053.2	11.00	6.0	10.0	
993	125	0	0	54.7	44.5						
994	125	-6	1	54.7	44.5	55.01	055.1	.70	.0	.0	
995	125	-5	2	54.7	44.5	55.01	055.1	.70	.0	.0	

Social scientists want to:

1) Produce and disseminate, and access other, Occupational Information Resources

2) Link together their (secure) micro-data with OIR's

External user (micro-social data)				Occ info (index file) (aggregate)				User's output (micro-social data)			
id	oug	sex	.	oug	CS-M	CS-F	EGP	id	oug	CS	
1	110	1	.	110	60	58	I	1	110	60	.
2	320	1	.	320	69	71	II	2	320	69	.
3	320	2	.	874	39	51	VIIa	3	320	71	.
4	874	1	.					4	874	39	.
5	874	2	.					5	874	51	.

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL   GEODE, 16 Jan 2007   NCeSS

We are agreed on how to do this:

✓ **Preservation of two levels of data**
  ▪ Index units: Occupational Unit groups, employment status
  ▪ Social classifications and other outputs
✓ **Use of transparent (published) methods** *[i.e. OIR's]*
  ▪ for classifying index units
  ▪ for translating index units into social classifications

for instance..

➢ Bechhofer, F. 1969. 'Occupations' in Stacey, M. (ed.) *Comparability in Social Research.* London: Heinemann.
➢ Jacoby, A. 1986. 'The Measurement of Social Class' *Proceedings from the Social Research Association seminar on "Measuring Employment Status and Social Class".* London: Social Research Association.
➢ Lambert, P.S. 2002. 'Handling Occupational Information'. *Building Research Capacity* 4: 9-12.
➢ Rose, D. and Pevalin, D.J. 2003. 'A Researcher's Guide to the National Statistics Socio-economic Classification'. London: Sage.

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL   GEODE, 16 Jan 2007   NCeSS

…but here come the buts...

Inconsistent preservation of source data
- **Alternative OUG schemes**
  - SOC-90; SOC-2000; ISCO; *SOC-90 (my special version)*
- Inconsistencies in other index factors
  - 'employment status'; supervisory status; number of employees
  - Individual or household; current job or career

Inconsistent exploitation of Occupational Information Resources
- **Numerous alternative occupational information files**
  - (time; country; format)
  - Substantive choices over social classifications
- Inconsistent translations to social classifications – 'by file or by fiat'
- **Dynamic updates** to occupational information resources
- **Strict security** constraints on users' micro-social survey data
- *Low uptake of existing occupational information resources*

GEODE, 16 Jan 2007

Two reactions and a proposed solution

1. **Enforce common standards**
   - In data collection and classification
   - E.g. Bechhofer 1969; Ganzeboom; Eurostat; ONS
     - *…on academic researchers..??!!*
2. **Give up**
   - *No attempt at engaging with published standards*

- Support plural occupational information resources in an accessible and consistent manner
  - **Internet facility coordinating OIR's**
  - **GEODE – Grid Enabled Occupational Data Environment**

GEODE, 16 Jan 2007

GEODE: Grid Enabled Occupational Data Environment

Objectives:

- ❖ **Create an international Virtual Organization for occupational data community**
  - Sharing, indexing, & curating diverse occupational data

- ❖ **Operate as a user-friendly portal**
  - Facilitate non-specialist user's access to occupational information
  - Search for and download occupational information
  - Support linkage from user's micro-data to OIR's

…and do this by exploiting 'e-Science' technologies..

GEODE, 16 Jan 2007

'The Grid' as a new technology

'The Grid' and 'eScience':
1. Online Coordination of electronic resources and collaborations
    - (Distributed computing)
    - Large scale
    - Collaborative
    - Heterogeneous
2. Standard protocols / information management systems

UK eSocial Science:
1) Investment in assessing / implementing technology
2) Computationally demanding data analysis
3) Qualitative and quantitative data collection technologies
4) ****Data sharing, processing and access****

GEODE, 16 Jan 2007

GEODE, eScience and eSocial Science

Some tentative comparisons...

	Similar to	*Diverges from*
GEODE-M metadata	DDI; Data Web; UKDA	[IDEAS]; [CAMSIS];
Data depository (OGSA-DAI)	ConvertGrid	[CASCOT]; [CAMSIS]; [ESDS]; Data Chronicles
Data matching service (OGSA-DAI + MDS)	[BRIDGES]	GEMEDA; Madiera
User engagement	Nesstar; ConvertGrid	CQeSS

GEODE, 16 Jan 2007

GEODE - architecture
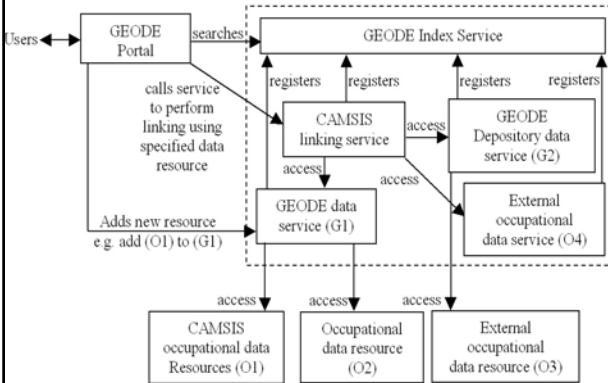
2) Occupational information depository

Storing occupational information resources

Considerations:
- All data stored at GEODE v's Linkage to external data
- Proprietary software *(plain text / SPSS / STATA)*
- Rectangular index files v's other formats (e.g. pdf)
  - 'index file' format favoured
- Finite number of occ info. files / model of **plurality** of supply
- International community of data providers
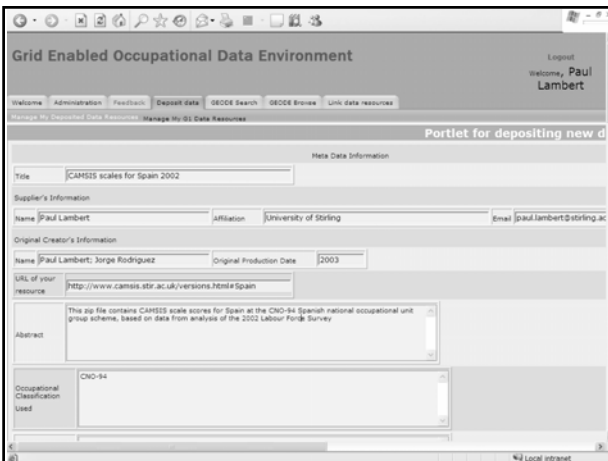- Negligible security restrictions (free online resources)

Strategy:
1) 'Uncurated' entry form, suits all formats, completed online
2) Curated entry (performed manually or automatically):
   - Translation to csv **index file**
   - Modify GEODE-M record for index file
- Storage: OGSA-DAI framework to link index files

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

GEODE, 16 Jan 2007

NCeSS

..compare with current practices..

3) Accessing and linking Occupational Information
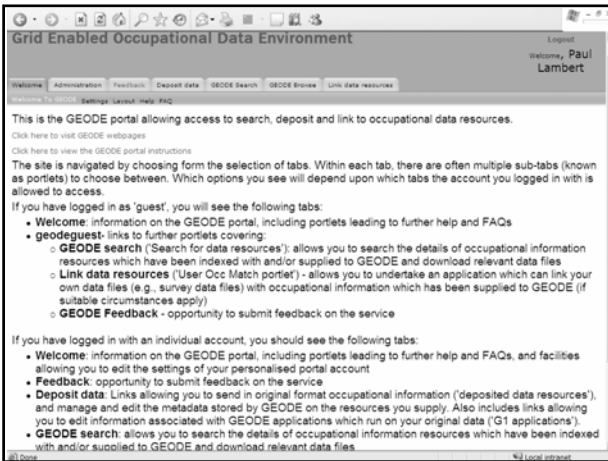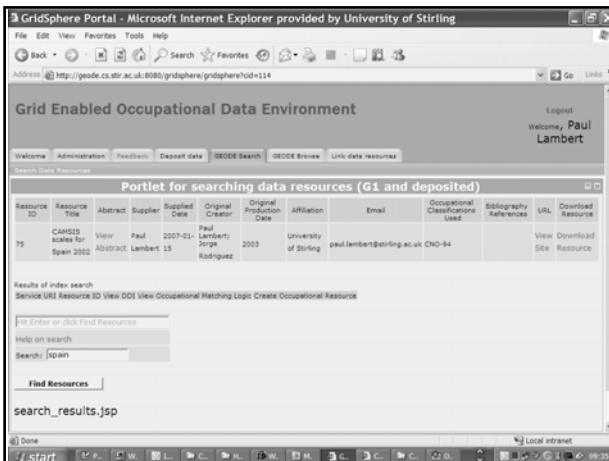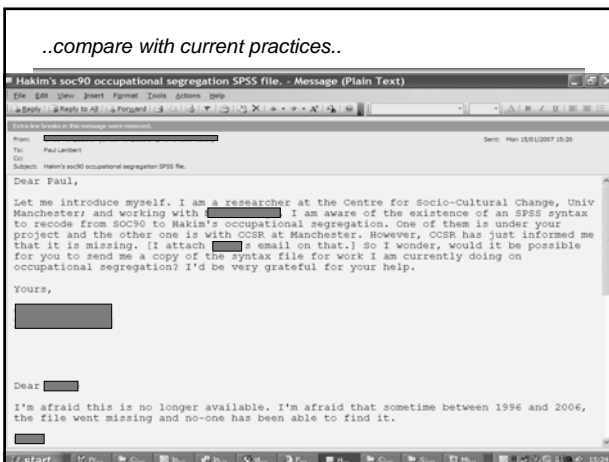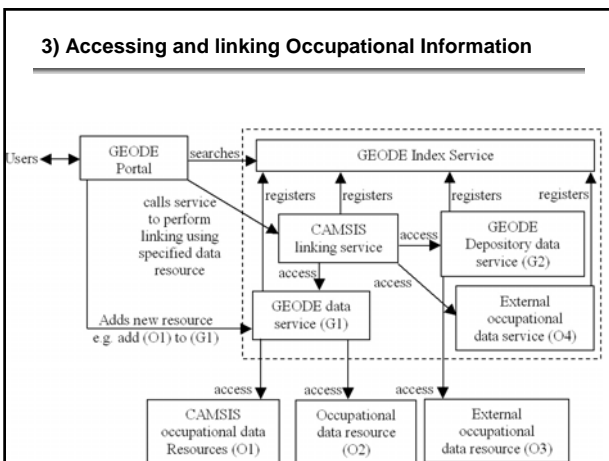
GEODE portal access

3.1) Searching and retrieving data

- GEODE 'search' and 'browse' facilities
  - Abstracts / descriptions
  - Time periods / countries / occupational units

- Further developments..
  - Improved search/browse algorithms
  - evaluative information ↔ GEODE data depositor's VO?

GEODE, 16 Jan 2007

Searching – uncurated resources

Searching – curated resources

GEODE portal access

3.2) File linkage mechanisms

Micro-social data (A) ↔ Occupational information resources (B)

- Multiple occupational variables on (A)
- Strict security constraints on (A)
- Inconsistent OUG formats on (A)

➢ JAVA application launched on users machine
  ❖ Simple file matching procedure
  ❖ Works on resources located at any URI
➢ Continuing development
  - Currently requires plain text input
  - Multiple occ. variables require repeated matching exercises
    (e.g. husband's occ.; wife's occ.)

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

GEODE, 16 Jan 2007

NCeSS

---
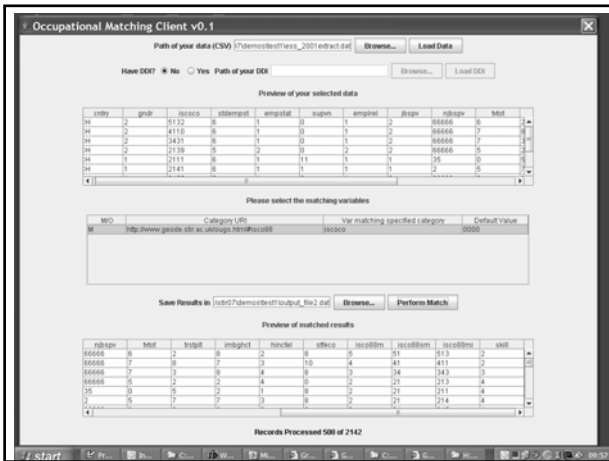
4) Summary – Handling Occupational Data

(1) Text records → OUG data

Currently:
Text coding software
  (e.g. CASCOT)
Manual look-up

GEODE:
Linkage to existing resources
Further facilities possible but not
  planned (users typically have
  adequate resources)

(2) OUG data → summary indicators

Currently:
Numerous aggregate **occupational information resources**
Bespoke data programming
  requirements

GEODE:
Core provision: management and
  access of these data resources
Service to large volumes of users

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

GEODE, 16 Jan 2007

NCeSS

GEODE Strategy - Conclusions and prospects

- **Occupational Information Depository**
  - OGSA-DAI implementations
  - Index-files annotated through 'GEODE-M'
    - Some ongoing manual support requirements

- **File linkage mechanisms**
  - JAVA application – still some user actions needed

- **Generic data service**
  - Hinges on numeric OUG index [cf. CASCOT]
  - other application areas – e.g. Education, Geography

- **Why use the Grid?**
  - E-Science project
  - Data linkage procedures
  - Data depository mechanisms (OGSA-DAI)
  - Extent of manual input?

GEODE, 16 Jan 2007

GEODE – user uptake

- High potential demand
  - Numerous queries on occupational data management
  - Numerous researchers wishing to distribute occupational data
- First GEODE services not yet user-friendly

- ☺ **Carrots**
  - High demands for easier access and review
- ☺ **Sticks**
  - Poor standards of many previous research which neglects good review of occupational information
- ☹ **Hurdles**
  - Change research cultures in social science disciplines(?)

GEODE, 16 Jan 2007