

GEODE: Grid Enabled Occupational Data Environment

GEODE Project introduction and summary, 12/12/05

• Motivation

- Many 'occupational information resources' in social sciences
 - Eg, class units, or status scores, for occupations
 - Vary by countries, time periods, updates – eg www.camsis.stir.ac.uk
- Social science survey data users wish to link their occupation data with such resources, but seldom manage successfully
 - Due to: Complex file management requirements; inconsistent existing information provision; vast quantity of resources; dynamic nature of resources

• GEODE collaboration (Oct 2005-March 2007)

- **Sociology** - Paul Lambert, Vernon Gayle (Stirling University); Ken Prandy (Cardiff)
- **Computing / eScience** - Larry Tan, Ken Turner (Stirling); Richard Sinnott (Glasgow)
- Online facility: to be accessed from www.geode.stir.ac.uk



GEODE Project introduction and summary, 12/12/05



What's the problem?

External user			Occ info (index file)				User's output		
id	occ	sex	occ	CS-M	CS-F	EGP	id	occ	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

But...

- Unreliable occupational data measurement
 - of National standards – ONS
- Inconsistent translations to social classifications
 - by file or by fiat
- Low uptake of occupational information resources
- Strict security constraints on users' micro-social survey data



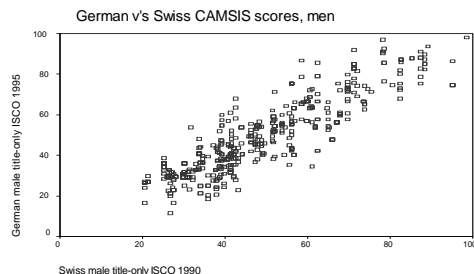
GEODE Project introduction and summary, 12/12/05



(1) Specificity and Universality

Theoretical, pragmatic and empirical issues

➤ Model for plurality in data supply



(2) Unit group schemes and sausage machines

EGP classifications: CASOC 1990 Vs ISCO88 (BHP/S)

	1 Service higher	2 Service lower	3 Routine non-mil employees	4 Personal service work	5 Sml props w employees	6 Sml props w/o employees	7 Farmers Smallhold ems	8 Foreman Technicians	9 Skilled manual workers	10 Semiskilled manual workers	11 Agricultural workers
8.00	13	10	0	0	0	0	0	13	0	0	0
1.00 I Higher Controllers (Service)	373	1116	819	83	20	18	7	0	70	0	2
2.00 II Low Controllers (Service)	171	131	707	117	22	11	5	1	38	0	2
3.00 III Routine NonManual	366	6	104	902	673	1	15	0	1	0	176
4.00 IVa Self-Emply with Employees	52	30	5	0	3	117	92	0	2	0	3
5.00 IVb Self-Emply no Employees	51	0	2	3	2	0	220	0	0	12	6
7.00 V Manual Supervisors	50	6	52	0	0	0	0	0	172	4	0
8.00 M Skilled Manual	139	0	38	1	0	2	10	0	41	476	70
9.00 VIIa Semi-Unskild Mil	313	11	33	33	44	3	45	0	243	87	1087
10.00 VIIb Farm Labour	14	0	2	0	0	0	5	5	0	0	23
11.00 IVc Selfempyl Farm	42	1	1	0	0	3	16	77	10	0	1



GEODE: Grid Enabled Occupational Data Environment

Objectives:

- Operate as a portal**
 - Facilitate linking occupational information to users' datasets
 - (initial focus on CAMSIS occupational information resources)
 - GEODE data resources – occupational information data curated as data service in Stirling, accessed by users via portal
- Create an international Virtual Organization for occupational data community**
 - Sharing, indexing, curating diverse occupational data resources
 - Aim to incorporate other analytical functions on occupational data

Uses:

- Globus Toolkit 4 (GT4)
- Grid Security mechanisms using PKI (Public-Key Infrastructure)
 - (external users' datasets may have stringent security conditions)

GEODE Project introduction and summary, 12/12/05



GEODE: Grid Enabled Occupational Data Environment

GEODE Data Resources

- Grid Data Service – uses OGSA-DAI middleware
- Curated data stored at Stirling
- Allows external users' data linkage via GEODE portal

Features:

- Need to deal with users' complex social survey datasets
 - (eg multiple occupational records, inconsistent formats, etc)
- Security restrictions on users' survey datasets
- High social science demand – many potential users
- Dynamic: data stores regularly updated

GEODE Project introduction and summary, 12/12/05

GEODE: Grid Enabled Occupational Data Environment

GEODE Virtual Organisation:

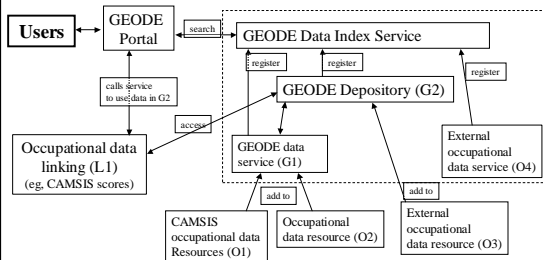
- Create a community for occupational data users (services and data resources)
- Allow service providers and data providers to add and publish their resources
- Features:
 - International community of social science researchers
 - Dynamic: regular additions / updates to shared resources



GEODE Project introduction and summary, 12/12/05



GEODE: Structure of Portal / Virtual Organisation



GEODE Portal (on behalf of users) searches the index for [CAMSIS] data; accesses the data from G2; and requests a [scale score] linkage via L1 (or requests another service using G2 as the data source)