Paul S. Lambert, Koon Leai Larry Tan, Kenneth J. Turner, Vernon Gayle, Ken Prandy and Richard O. Sinnott. Development of a Grid Enabled Occupational Data Environment. In Rob Procter, editor, Proc. 2nd International Conference on e-Social Science, pages 1-12, National Centre for e-Social Science, Manchester, UK, June 2006.

Development of a Grid Enabled Occupational Data Environment

Lambert, P.S.[1], Tan, K.L.L.[1,2], Turner, K.J. [2], Gayle, V. [1], Prandy, K. [3], Sinnott, R.O.[4].

[1]University of Stirling, Department of Applied Social Science, UK

[2]University of Stirling, Department of Computing Science and Mathematics, UK

[3]Cardiff University, School of Social Sciences, UK

[4]University of Glasgow, National e-Science Centre, UK

Email address of corresponding author: paul.lambert@stirling.ac.uk

Abstract. The GEODE project is developing user-oriented Grid-based services, accessible via a portal, for social scientists who require and use 'occupational information' within their research. There are many complexities associated with social scientists' use of data on individual occupations. These arise for example from the availability of numerous alternative occupational classifications, and the use of different occupational definitions across countries. This paper describes how the GEODE project is developing an online service which acts as a facility supporting access to numerous occupational information resources. This is achieved through an integrated Grid service which uses a Globus Toolkit 4 infrastructure and OGSA-DAI (Database Access and Integration) middleware to provide the necessary data indexing and matching services, accessed through a user-oriented front-end portal (using GridSphere). The paper discusses issues in the implementation and organization of these services.

1. Introduction

1.1 Occupational Information

In this paper we describe how the GEODE project (Lambert *et al.* 2006) is used to facilitate the provision of occupational information to social researchers. The focus is upon providing facilities for the collation and distribution of aggregate occupational information. Although such data is relevant to many social research disciplines, previous uses of occupational information have flaws which may contribute to the under-use, or even misuse, of a range of valuable information resources (Bardasi 2004; Lambert 2002; Blackwell 2001).

There are broadly two relevant forms of occupational information. The first involves the measurement of data on the occupational titles of individuals (for instance, typically collected during social survey questionnaires). Such data is usually coded into a standardized scheme of occupational definitions (known as Occupational Unit Groups, OUGs). There are several alternative OUG schemes, and existing facilities to support the coding of occupational descriptions into appropriate OUG units (e.g. Jones 2006; ONS 2005b; van Leeuwen *et al.* 2005; ILO 1990). The GEODE project does not cater extensively to this type of occupational

information coding, and users of GEODE are usually expected to have linked their resources with some form of OUG system (a methodological best-practice which is, nevertheless, not universally adopted by social researchers, cf. Lambert 2002).

The second form of occupational information involves aggregate data about particular OUG units (or aggregations of OUG units). Examples include data on the average earnings of particular occupations in contemporary Britain (e.g. McKnight and Elias 1997), or the average proportion of women in different OUGs (e.g. Hakim 1998). These resources also incorporate 'translation' or 'conversion' tools, which are used to locate OUG units into a substantively meaningful occupationally based social classification scheme, such as a class scheme (e.g. ONS 2005a), or a stratification scale (a score assigned to OUGs, indicating relative positions of advantage or disadvantage, e.g. Ganzeboom 2006). In fact these coding resources are in high demand amongst social researchers, since a common motivation for collecting occupational data is to use it to derive an occupationally based social classification measure. There is a substantial literature focusing upon evaluations of alternative occupationally based social classifications (e.g. Rose and Pevalin 2003), but there has been far less attention paid to the practical use of alternative schemes.

There have been previous Internet and manual provisions of aggregate occupational information to social researchers, including the CAMSIS project (Lambert and Prandy 2006), the data from which is used in the GEODE project prototypes described below. Table 1 lists a selection of existing aggregate occupational information resources which are amongst those being incorporated within the GEODE service (many further resources are also used). As summarized in Table 1, most aggregate occupational information resources comprise relatively small electronic files, for instance, a matrix or translation index ranging over a few hundred different OUG units.

Several problems have arisen for social researchers attempting to distribute and utilize such resources. Firstly, the existing facilities available for those researchers who wish to access such data are uneven. Data is often supplied in a limited range of electronic formats (data files supplied in the format of the proprietary software SPSS are the most common). The data tend to impose relatively demanding programming requirements on the user. For instance, many users wish to connect other micro-social data – featuring an OUG code - with aggregate occupational information, but to do this they are typically asked to download a translation file then undertake some bespoke programming in order to process this data. They often cover only subsets of appropriate occupations (for example OUG differences at a limited level of aggregation). Moreover, these inconsistencies have been exacerbated as the variety of available aggregate occupational information has expanded dramatically – Table 1 indicates that whilst the size of aggregate data files is small, their volume is substantial.

Additionally, facilities for those researchers wishing to supply and distribute aggregate occupational information to others have tended to be uncoordinated, lacking common standards for metadata and semantic annotations; for file formats; and lacking data quality evaluation. In particular, aggregate occupational information is often published through web links with little or no documentation except for natural language instructions, which can be ambiguous, within accompanying materials. Many data resources are also not indexed with other services and therefore do not experience good exposure within the research community.

A consequence has been methodological difficulties for contemporary social researchers (Lambert 2002). Whilst substantive interest in alternative occupational information resources is high, and the availability of occupational data is widespread, relatively few social researchers have been comfortable in accessing and processing appropriate aggregate

occupational information. The problems are primarily concerned with technical limitations. Below we describe how the GEODE project uses Grid computing facilities to address these problems.

Table 1: Aggregate Occupational Information Resources as used in GEODE service

Format	OUGs / index units [average no. distinct units]	Output	Est. no. distinct files	Typical file size
1.1 CAMSIS indexes, www.camsis.stir.ac.uk/versions.html				
Index matrix, SPSS and plain text data files	Variety of national OUGs; gender; employment status *[2000]*	CAMSIS scale scores for UG/es/ gender combination	200	100kb
1.2 CAMSIS occupational title value labels, www.camsis.stir.ac.uk/occunits/distribution.html				
One-to-one translation, SPSS syntax and plain text files	Variety of national OUGs *[400]*	Text labels for numeric OUG codes	50	50kb
2.1 ISEI tools, home.fsw.vu.nl/~ganzeboom/pisa/				
One-to-one translation, SPSS syntax and plain text files	ISCO-88 and –68 OUG schemes *[533]*	ISEI and SIOPS occupational scale scores	20	50kb
2.2 ISEI tools at IDEAS-REPEC, ideas.repec.org/c/boc/bocode/s425802.html				
One-to-one translation, STATA do files	ISCO-88 and –68 OUG schemes *[533]*	ISEI and SIOPS occupational scale scores	20	50kb
3.1 E-SEC matrices, www.iser.essex.ac.uk/esec/				
Index matrix, MS-Excel and SPSS syntax	ISCO-88 OUG; employment status (es) *[4000]*	E-SEC class position for OUG-es combination	20	200kb
3.2 NS-SEC translation matrix, www.statistics.gov.uk/methods_quality/ns_sec/				
Index matrix, MS-Excel and plain text	UK SOC90 and SOC 2000; employment status *[3000]*	NS-SEC class position for OUG-es combination	6	200kb
4. Hakim gender segregation codes (Hakim 1998)				
One-to-one translation, paper publication	ISCO-88 and UK SOC90 *[400]*	Gender segregation information for OUG code	2	*paper*
5. HISCO occupational labels and codes, http://historyofwork.iisg.nl/				
One-to-one translation, SPSS syntax and plain text files	HISCO; country of study *[500]*	HISCO OUG unit descriptions, by country	5	50kb

Notes:

- 'Syntax' and 'do' files are plain text command files for SPSS and STATA software respectively.
- The 'est. no. distinct units' refers to the typical number of index units associated with any one component file, where an index unit comprises a unique combination of OUG and any other contributing factors

1.2 GEODE project objectives

The Grid defines a scalable architecture where data and computational resources are virtualised, abstracted, and collaborated on within virtual organisations (Foster *et al*. 2001). Both the distribution and utilisation of occupational information resources are amenable to a Grid architecture. The GEODE project aims to improve the current utilisation of occupational

information by using the Grid. The goal is to create a virtual community where occupational information resources are virtualised, indexed, and are uniformly accessed by users in a secure manner resulting in a gateway where occupational information can be discovered, exchanged and collaborated on. The long-term objective is to improve standards in the use of occupational information by social researchers.

Figure 1 gives a schematic illustration of the architecture used by the GEODE project to achieve these objectives. Most of the tasks incorporated in the project can be located in one of two areas. The first concerns a coordinating index service illustrated by the systems included within the dotted box of Figure 1 (see 1.2.1). The second concerns the upper left segment of Figure 1, the provision of a linking and analysis service to external users with little previous contact with aggregate occupational information resources (see 1.2.2).

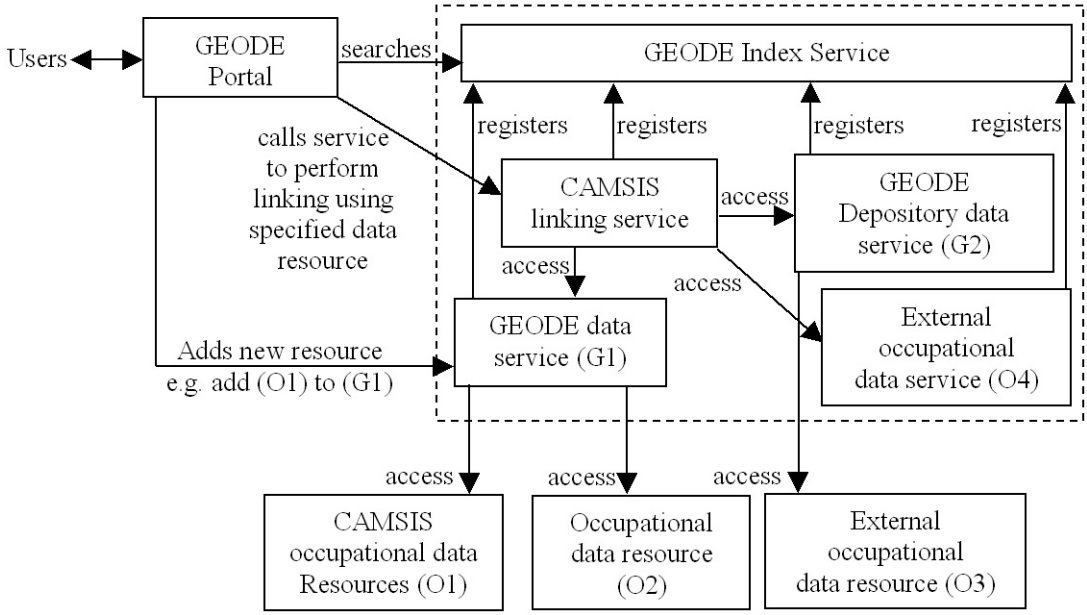

Figure 1. Architecture of GEODE.

1.2.1 Coordinating index service

The GEODE project seeks to provide a service known as the 'Occupational Information Depository' which facilities the indexing of diverse forms of aggregate occupational information such as described in Table 1. The motivation for this service arises from the limitations of current practices in occupational information distribution. In Figure 1, the resources marked O1, O2, O3 and O4 are used to indicate examples of occupational datasets. The GEODE service uses a customized metadata system (see 2.2) to achieve a common standard in semantic occupational data description. Such metadata may be registered in the GEODE index service. The different linkages shown in Figure 1 between the 'GEODE index service' and the original occupational information resources are used to depict alternative arrangements for the derivation and indexing of occupational information.

1.2.2 Portal linkage and analysis service

The GEODE project also seeks to develop a highly accessible application which will enable non-specialist social scientists to readily access the aggregate occupational information indexed by the GEODE service. In Figure 1, this application is indicated by the links from the 'GEODE Portal', a web interface access to the GEODE occupational

information resources. In some circumstances, this access may involve analytical requests to summarize the aggregate occupational information resources. Most often however, the access will involve researchers using the portal service in order to link their own micro-social datasets (which contain OUG records) with appropriate aggregate occupational information.

We are not aware of any previous service of such a nature in the social sciences. Instead social researchers are generally required to undertake a series of programming activities themselves in order to achieve these types of task. Occupational information resources make a particularly interesting illustration of this matching process, because of the combination of the widespread use of occupational information resources; the relatively low levels of data management and processing skills held by the user community; and the high volume, and lack of coordination, which characterizes the supply of aggregate occupational information. In section 3, we describe the issues arising in achieving an accessible portal service that addresses these circumstances.

2. Occupational Information Depository

2.1 Metadata standards

Previous provisions of aggregate occupational information have not used consistently defined metadata. Such data is essential to indexing information within a Grid enabled environment (cf. Cole *et al.* 2003; NCeSS 2006), and also brings benefits for data access and search capacities. Given that most occupational information data providers have limited previous exposure to formalized standards of metadata documentation, the task for the GEODE project has been to identify an appropriate metadata system which encompasses information important to occupational systems; is compatible with requirements for Grid enabling facilities; and can be presented in a manner accessible to social science data curators.

2.2 GEODE-M: Semantic metadata formulation

Formalized XML schemes organize the appropriate content to allow for Grid enabling datasets (cf. Cole *et al*. 2003). The GEODE service uses a customized XML metadata system that is based upon the Michigan Data Documentation Initiative (DDI 2005). The DDI system was chosen as it offers considerable flexibility in metadata definitions and is widely established across research disciplines (Cole *et al*. 2003).

The appropriate customisation should balance expert judgments on the information necessary to document resources, with practical evaluations of working implementations of the proposed framework. Figure 2 gives an illustration of the contents of the GEODE recommended XML metadata customisation (‘GEODE-M’), the final version of which is likely to be adjusted pending future activities.

The items listed as elements in Figure 2 refer to information concerning the occupational information resources recorded within the GEODE-M scheme by means of a specific XML tag. This customized metadata standard is used to provide a resource for drawing out the salient identifying features of aggregate occupational information datasets such as described in Table 1. The standard is devised in such a way as to minimize the requirements for describing occupational data (under the expectation that data suppliers may be able to curate their own data). The current standard is undergoing an evaluation process, which involves checking that sufficient information is incorporated to adequately describe the various

resources identified by the GEODE project researchers (such as those listed in Table 1). The organisation of the DDI system is premised upon the 5 'structures' illustrated in Figure 2: these prove convenient since the <dataDscr> structure incorporates the aspects of (multiple) variable definitions common to aggregate occupational information resources, whilst the other structures span a range of topics suited to defining clear information about occupational information resources.

Figure 2. Structure of GEODE-M (DDI-based recommended metadata for curating aggregate occupational information resources)

The core feature of the standard is the flexible specification of identifier variables (a relevant OUG value, and any other identifying factors). This is important because different aggregate occupational information resources use different combinations of identifiers (see Table 1). In addition, the standard requests clear definitions of the output data supplied by the

aggregate information resources, and the provenance, release date and context of the resources. The term 'context' is used as shorthand to refer to the social context of the data, such as the country or time period to which it is applicable. Different contexts are a principle reason why the number of distinct files associated with an aggregate occupational information resource often proliferate (Table 1), as a different file is usually used for each context. However in some provisions, 'contexts' may be indicated through an 'identifier' variable rather than a separate data file, depending on the preferences of the data suppliers.

As Figure 1 illustrates, the activities of the GEODE project include the provision of some metadata for existing data files being undertaken by members of the project (i.e. G1). However the project also seeks a framework allowing external users to undertake a semi-automated data curation process (e.g. G2 in Figure 1). The latter service is attractive because of the large volume of distinct aggregate occupational information resources, and the dynamic nature of such resources (several of the resources listed in Table 1 have been updated over time, despite retaining the same structure and functionality). By keeping the GEODE-M proposed standard as simple as possible, it is intended that a web-based proforma can be designed to enable rapid completion of necessary metadata via the GEODE portal service.

2.3 Data file contents

The ability to make use of such metadata hinges upon the appropriate formatting of the underlying data file. It is necessary that the aggregate occupational information is supplied in a fashion whereby the OUG and any other identifier variables may be easily recognized via the metadata. This requires that the data be supplied in a format that is accessible to the GEODE index service. Lord and Macdonald (2003) highlight the dangers of data supply via proprietary software, but, as may be seen from Table 1, many occupational information files are supplied in such formats. Although one possibility might be to restrict GEODE services to formats which are not software specific, this would represent a significant conflict with the interests of social science researchers, who work overwhelmingly within a selected range of proprietary packages. To some extent, the dominance of two particular proprietary packages over the last decade within the social survey research field (SPSS, www.spss.com, and STATA, www.stata.com), provides the opportunity for an intermediate solution. The GEODE project uses local licences to these packages to permit the reading, and later the supply, of both formats of data within its service (although the underlying data processes are conducted using plain text formats). In practice, catering to these software packages, and to plain text files, has allowed access to all external occupational information files currently under review, although there is a clear possibility of future conflicts with this solution.

It is thus envisaged that most occupational information data depositors will need merely to supply an additional metadata component to the materials that they release, permitting them otherwise to retain the same formats as they currently release. It could be argued that this represents a substantial improvement over existing services. A few alternative resources do offer to distribute other occupational information (cf. IDEAS 2006). However these make requirements that any data supplied must be processed and documented to a specific format.

2.4 Collaborative Virtual Community

The community of occupational information data providers is large and international. As such the provision of the Occupational Information Depository service acts to support a collaborative virtual community of occupational information suppliers. Grid security protocols provide a convenient structure for this organisation. The GEODE project uses the

MDS (Monitoring and Discovery System, e.g. Schopf 2006), provided by the Globus Toolkit, which provides data aggregation and notification services. The organisation allows for fine-grained control over user access and the security of the data resources. This enables an indexing service to be deployed to hold registry information on resources and services and subsequently enable the sharing of data. Resources register with the organisation through the indexing service.

2.5 Occupational Information Depository - GEODE Prototype

A prototype development of the GEODE services for virtualising and distributing occupational information resources has been established. The GEODE infrastructure leverages the data abstraction Grid middleware (OGSA-DAI 2006) to create a framework for dynamically deploying occupational information data resources. The OGSA-DAI middleware, in addition to being able to automatically perform registration with the indexing service, contains the provisions to register custom metadata together with the database schemas. OGSA-DAI also provides a framework for automatic derivation of data access metadata using the GEODE-M scheme, with provisions made available for custom metadata if required.

Based on OGSA-DAI configurable data services and appropriate drivers, the GEODE data services allow for the data abstraction of aggregate occupational information. This helps to address problems of heterogeneity of data sets and confusion over the availability and applicability of alternative resources. The GEODE prototype service follows the organisation of Figure 1. GEODE maintains two data services: G1 virtualises data that is curated locally; and G2 is a collection of resources harnessed from a wider international community of social scientists. O1, O2 and O3 are examples of the data resources abstracted by the respective data services. There can also be an arbitrary number of external data services like O4 registering in the GEODE grid, but for illustration purposes only one (O4) is shown.

3. Access to Occupational Information

3.1 Portal configuration

Many social science researchers are unfamiliar with advanced computer applications. Therefore it is desirable to develop GEODE as an application that can be used with minimal learning and configuration. Though a custom application has been considered, a web portal is much more appealing to the users. GEODE has developed a web portal as the user interface by which occupational data researchers can interact with the grid infrastructure and the resources stored within the Occupational Information Depository. The portal caters both to depositors of data resources (through which users can administer their data resources), and external users wishing to access aggregate occupational information by searching the data index and making requests for statistical services. The portal is accessible via the Internet using standard web browsers. Application users are not bound to specific machines and software in order to perform tasks. This will greatly increase usability in the social science community.

The portal was developed with the GridSphere Portal Framework (2005), an open-source and widely used tool for portal development. GridSphere is used to develop the components that make up the portal, namely the presentation view, presentation logic, and the application logic. The view is implemented with JSP and the logic with portlets that control the presentation flow. The emphasis is on the application logic, developed as a portlet service,

which interfaces with the Grid environment. The portlet service invokes the operations of the Grid services and returns results to the presentation logic. GEODE follows the Model-View-Controller design pattern which the occupational data Grid (model), presentation (view) and portlet service (controller) represent.

3.2 GEODE data services

The Grid-specific services are built with Globus Toolkit 4 (GT4, see www.globus.org). This implementation is based on an OASIS standard (WSRF 2006). At present, GEODE has developed specific services to make queries to the index service, and to link occupational data to CAMSIS scale scores. As the scope of GEODE evolves, further services can be readily implemented and deployed on the Grid and accessed via the portal. The application services provide specific data analysis functions on the data resources.

3.2.1 File Linking mechanisms

The task is to provide an easy to use facility which allows owners of micro-social data (for instance survey data) to declare which parts of their data link to the aggregate occupational information, and then to process this linkage. In (2.4), we noted that security protocols were used to assist in the organization of the Occupational Information Depository. In practice there are few security concerns concerned with aggregate occupational information. However, there are substantial security restrictions over access to and linkage with social science micro-data. Ordinarily, owners of micro-social datasets are bound by agreements which prevent them from releasing the data from a secure environment (although some resources are less rigidly constrained). This provides a key challenge for the GEODE linkage service.

A proposed solution involves providing facilities which allow users first to declare the subset of micro-social data which features the occupational information identifiers (OUG variables and in some circumstances other context variables). An application is then deployed which temporarily separates the occupational information identifiers, and merges them with the relevant aggregate occupational information, before re-connecting to the source data. This provision would require the implementation of X.509 digital certificates to facilitate the differential access levels associated with the stages of the process. Globus Toolkit may be used via its implementation of GSI (2006) to establish basis security in a Grid environment. GSI offers authentication, authorization, credential delegation, and single sign-on that GEODE could leverage to administer resources and portal service operations. OGSA-DAI makes it possible for resource security to be configured when deployed. Users delegate their credentials (proxy certificates) to the GEODE services to allow operations to be carried out on their behalf and accounted for.

A requirements analysis is ongoing for alternative implementations of this solution across the variety of data requirements and interests of social research micro-data users (see also section 4.2). The priority is to minimise the level of interactive activity requested of the micro-data holders, which ideally may involve no overt certification registration and no application installation (many potential users of the service, for instance, may not have administrative access to the computers on which the work).

3.2.2 Analytical queries

Another service under development within the GEODE portal involves facilities for immediate review and analysis of the occupational information resources indexed under the

Occupational Information Depository. Several analysis functions are provided under the prototype service, including a data search facility and the possibility of extracting summary statistical data on the properties of different occupational units. In such applications users interact with the data services via application services which in turn are invoked through the portal. These developments offer exciting opportunities to social science researchers, who hitherto would be required to trawl across a vast terrain of occupational information resources and publications in order to retrieve relevant data.

4. Conclusions and Prospects

4.1 Project Progress

DDI has been incorporated manually and tested successfully in registering and retrieving information from the indexing service. GEODE is currently developing the portal interface and data service activities to manage the proposed 'GEODE-M' metadata definitions interactively. The requirements for linking micro-social data resources to aggregate occupational information are currently being developed and evaluated.

GSI and credential delegation will be assessed and implemented once the functional requirements of GEODE are finalised. To simplify the complexities of client GSI set up, creating proxy certificates and delegating credentials, Java WebStart (2006) is being investigated. This allows applications to be deployed and launched with a single click from a Web browser, thus omitting complicated and specific installations for GEODE users. In addition, this allows researchers to utilise GEODE on other machines instead of being constrained to their own machines. Also, updates to the services and datasets would automatically be available to the users through this WebStart approach.

4.2 Lessons learnt

OGSA-DAI supports activities which describe and carry out sequences of operations on data. Currently however these activities are implemented such that they are all invoked via one single operation flow. Therefore it is not possible to support the different security configurations that GT4 supports for individual operations. Although not critical presently, it may become a growing consideration that will impact the project practically. For example an activity to modify the resource metadata may have a requirement of authorisation using an access control list, which itself requires activity-level security configuration that OGSA-DAI currently does not support.

There are a few ways to delegate credentials to services within GT4. One involves a client performing the delegation to the services directly. Alternatively the client can store credentials in a depository, where services are then informed of details to retrieve the credentials in order to have the delegated rights. The latter is more favourable as it does not confine proxy credentials to specific locations to perform delegation.

There are currently two implementations of credential delegation, namely MyProxy and GT4's DelegationService (Globus Alliance, 2006a and 2006b). MyProxy is only available as a software installation in Unix/Linux. GEODE is implemented under Windows and it is preferable to keep a single environment for maintainability. The DelegationService is a Grid service that provides similar functions. Hence this service can be easily deployed into Grid service containers. GEODE aims to use the DelegationService as the proxy credential depository. However the current limitation in using the DelegationService is that credentials

can only be delegated to services that run within the same service container. In cases where services are deployed in multiple containers, this limitation may be resolved by having a DelegationService deployed in each container. Ideally the DelegationService should be used to perform delegation to services in disparate containers.

4.3 Future developments

The GEODE occupational information depository architecture is being designed to be scalable. An arbitrary number of data services can be deployed (possibly on different machines/sites) to handle the vast amount of virtualised occupational information datasets, both internal and external in origins. The primary development intended for the GEODE service involves expanding the scope of occupational data resources and reviewing the possibilities for their integration with other external datasets.

The GEODE architecture is designed to be as generic as possible. One of the most promising benefits is to be able to apply or extend the structure towards other social science statistical data resources with similar requirements. In a generic context, a data Grid with registration to index services along with implemented services can fulfill the requirements of data sharing and collaboration to a substantial degree. In addition to data abstraction and location transparency, this architecture allows control of services whereby the data provider may have the flexibility to provide data services as well as using the services set up within GEODE. Thus in principle the GEODE Grid can be extended and used for non-occupational social science data. For example different DDI metadata customisations could be prepared for alternative data resources in instances when social scientists have similar requirements for both the storage and distribution of data. Possible areas of application may include the management of geographical and educational data resources, although the scope of this project does not include such implementations.

The key objectives of the GEODE project are to address problems of occupational information data access and distribution. The facilities described offer exciting opportunities to do this, although the key test rests on the planed evaluations of user uptake for prototype and final resources. These will also offer insights into another aspect of the project development, the extent to which social scientists can make comfortable use of the Grid services embodied in the GEODE architecture.

Acknowledgements

This research is supported by the ESRC's 'Small Grants in eSocial Science' programme, reference RES-149-25-1015.

References

Bardasi, E. (2004) "The Luxembourg Income Study and the Luxembourg Employment Study as resources for labour market research" *Socio-Economic Review*, 2(4)239-261.

Blackwell, L. (2001) "Women's work in UK official statistics and the 1980 reclassification of occupations". *Journal of the Royal Statistical Society, Series A* 164, 307-325.

Cole, K., Schurer, K., Beedham, H., Hewitt, T. (2003) *Grid Enabling Quantitative Social Science Datasets – A Scoping Study.* Swindon: ESRC / JISC scoping study, and http://www.ncess.ac.uk/docs/grid_enabling_quant_data.pdf

DDI (Data Documentation Initiative) (2005) http://www.icpsr.umich.edu/DDI/ [15/11/2005]

Foster, I., Kesselman, C., and Tuecke, S. (2001) The Anatomy of the Grid: Enabling Scalable Virtual Organizations. *International Journal of Supercomputer Applications*, 15(3).

Ganzeboom H.B.G. (2006) *Tools for Deriving Status Measures from ISK0-88 and ISCO-68.* http://home.fsw.vu.nl/~ganzeboom/PISA/ Amsterdam: Amsterdam University.
Globus Alliance (2006a) *The Globus Alliance, GT4.0: Credential Management: MyProxy*, http://www. globus.org/toolkit/docs/4.0/security/myproxy/, [4/2006].
Globus Alliance (2006b) *The Globus Alliance, GT4.0: Security: Delegation Service*, http://www. globus.org/toolkit/docs/4.0/security/delegation/ [4/2006].
Gridsphere Portal Framework (2005), http://www.gridsphere.org [12/2005]
GSI (2006) The Globus Alliance, Grid Security Infrastructure, http://www.globus.org/toolkit/docs/4.0/security/key-index.html [4/2006].
Hakim, C. (1998) *Social change and innovation in the labour market.* Oxford: Oxford University Press.
IDEAS (2006) *IDEAS: Economics and Finance Research.* http://ideas.repec.org/ [4/5/2006]
ILO (1990) *ISCO-88: International Standard Classification of Occupations.* New York: International Labour Office.
Java WebStart (2006) Java WebStart (CoG Kit), http://www.cogkit.org/ [2/2006]
Jones, R. (2005) CASCOT project webpages, http://www2.warwick.ac.uk/fac/soc/ier/publications/software/cascot/ [19/12/2005]
Lambert, P.S. (2002) "Handling Occupational Information". *Building Research Capacity.* 4,9-12.
Lambert, P.S. and Prandy, K. (2006) CAMSIS project webpages, http://www.camsis.stir.ac.uk/ Stirling: University of Stirling. [24/3/2006]
Lambert, P.S., Tan, K.L., Turner, K., Gayle, V., Sinnott, R., Prandy, K. (2006) Grid Enabled Occupational Data Environment, http://www.geode.stir.ac.uk/ Stirling: University of Stirling. [8/5/2006]
Lord, P. and Macdonald, A. (2003) *Data curation for e-Science in the UK.* The Digital Archiving Consultancy Limited (e-Science Curation Report prepared for the JISC Committee for the Support of Research).
McKnight, A. and Elias, P. (1997) "A database of information on unit groups of the Standard Occupational Classification". In: Rose, D. and O'Reilly, K., (Eds.) *Constructing Classes.* Colchester: University of Essex.
NCeSS (National Centre for eSocial Science) (2006) Insight: Case Studies: Grid Enabling Quantitative Datasets Case Study Series, http://www.ncess.ac.uk/insight/case_studies/grid_en/ [10/5/2006]
OGSA-DAI (2006), Open Grid Service Architecture, Data Access and Integration, http://www.ogsadai.org.uk [2/2006].
ONS (2005a) *The National Statistics Socio-Economic Classification -. http://www.statistics.gov.uk/methods_quality/ns_sec/* London: Office for National Statistics. [20/7/2005]
ONS (2005b) *Standard Occupational Classification 2000 - http://www.statistics.gov.uk/methods_quality/ns_sec/soc2000.asp* . London: Office for National Statistics. [10/5/2005]
Rose, D. and Pevalin, D.J. (2003) *A Researcher's Guide to the National Statistics Socio-economic Classification.* London: Sage.
Schopf, J.M. *et al*. (2006) *Monitoring and Discovery in a Web Services Framework: Functionality and Performance of the Globus Toolkit's MDS4*, Argonne National Laboratory Technical Report #ANL/MCS-P1248-0405, and http://www-unix.mcs.anl.gov/~schopf/Pubs/mds-sc05.pdf .
van Leeuwen, M. *et al*. (2005) *History of Work Information System, http://historyofwork.iisg.nl/* . Amsterdam: International Institute for Social History [2005]
WSRF Web Services Resource Framework (2006) WSRF Specification, v1.2, http://www.globus.org/wsrf/ [3/4/2006].