ConvertGrid: Grid enabling existing social science data infrastructures and services.

Keith Cole[1], Pascal Ekin[2], Linda Mason[1], Jon Maclaren[3]

[1]MIMAS, Manchester Computing, University of Manchester, UK

[2]RSS, Manchester Computing, University of Manchester, UK

[3]Center for Computation and Technology, Louisiana State University, USA

keith.cole@manchester.ac.uk

Abstract. The aim of the ConvertGrid project was to demonstrate the advantages that Grid technologies can bring to the social sciences, by using such technologies to address key problems facing researchers who want to combine data from multiple geo-referenced data sets. The ConvertGrid project has developed a Grid based solution to this problem of geographical conversion. The ConvertGrid system has expanded upon the capabilities of the existing Convert service by linking it to potential data resources via the Grid. It enables users to select data of interest across a range of topics from multiple datasets, convert this data to a common target geography using geographical conversion tables. The resultant data streams are combined and returned to the user or transferred automatically to a web based mapping/visualisation interface. All the data retrieval and geographical conversion processes are hidden from the user. The ConvertGrid project has demonstrated how Grid technologies can be used to automate very complex workflows and also help to stimulate novel forms of research through promoting increased and more effective use of multiple data sources. However, developing a Grid based service to access, integrate and analyse multiple datasets which is simple to use has required the project team to address a number of key technical and/or methodological challenges.

1. Background

The ConvertGrid project was funded as part of the ESRC Pilot Projects in e-Social Science Programme. The overall aim of the ConvertGrid project was to demonstrate the advantages that Grid technologies can bring to the social sciences (Cole, Schurer, Beedham and Hewitt 2003), by using such technologies to address two common problems facing researchers who want to combine data from multiple geo-referenced data sets. The problems in question are:

- the initial *data management problem* of geography conversion, and
- the subsequent *data fusion problem* of combining the converted data.

These problems are best demonstrated by the following simple example. A researcher with postcoded GP patient health data wishing to study the relationships between the recorded health patterns and local conditions at Primary Care Trust (PCT) level might require access to a range of indicators covering a variety of different topics but these indicators are only

available from a number of different data sources, each with its own native geography (e.g. the 1991 Census LBS/SAS (1991 Census geography); the Experian Postcode Sector dataset (2000 postcode sector geography) and the ONS Neighbourhood Statistics (1998 local authority and ward geography)).

The researcher would have to extract the required data items from each dataset via a different native interface and convert them to the PCT target geography. Once data was extracted from each source in the native geography of the dataset the existing Convert service (see section 3.1) could be used to convert it to the target PCT geography (without the Convert service the process would be virtually impossible for all but the most hardened specialist with access to and extensive knowledge of the All Fields Postcode Directory). Following this, the data would have to be combined, before finally being ready for analysis. Whilst this is a simple example, there is a whole range of substantive research topics which are constrained because of the problems relating to geographical conversion.

The ConvertGrid service will expand upon the capabilities of the existing Convert service by linking it to potential data resources via the Grid. This will offer users the ability to select data of interest from multiple datasets, convert this data to a common target geography, and then combine this data in a single operation with the resultant data streams returned to the user or transferred automatically to a web based visualisation interface. All data retrieval and conversion will be hidden from the user. This will have the effect of improving researchers' productivity by automating very complex workflows as well as increasing use of multiple data sources by less technically competent users.

2. Project Objectives

The overall aim of the ConvertGrid project was to demonstrate the advantages that Grid technologies can bring to the Social Sciences. The specific objectives of the ConvertGrid project were:

1) to provide a practical demonstration of how Grid technologies can be used to facilitate the integration of data from disparate sources and overcome a major barrier to research use of multiple datasets;

2) to demonstrate how it is possible to Grid-enable existing socio-economic data sources and associated data services and make them available to other Grid-based services;

3) to demonstrate the potential of the Grid for teaching purposes by building a user interface to a Grid-based service which is suitable for student use;

4) to demonstrate how Grid technologies can be used to extend the functionality of existing data services;

5) to illustrate how the integration of multiple datasets from different sources can enhance the analysis/visualisation of spatially referenced socio-economic datasets.

All of these objectives were achieved.

3. Approach

3.1 Utilising the current Convert system

One of the objectives of the project was to use Grid technologies to extend the functionality of the existing Convert system. The Convert system is the result of an ESRC funded project (award H507255164) which involved the derivation and validation of geographical lookup tables from the 1999b version of the All Fields Postcode Directory (AFPD). The Convert website, which is an operational service hosted by MIMAS at (http://convert.mimas.ac.uk/), provides an accessible interface to the underlying system, allowing users to make use of the lookup tables to convert their own input data between many of the 25 administrative, electoral, census and postal geographies held on the AFPD. Users are also able to download the underlying lookup tables in their entirety or as geographical subsets. Since its launch in 2001, the Convert website has been heavily used, and has generated a great deal of interest and enthusiasm amongst academic and non-academic users.

3.2 Data selection and use cases

For the purpose of this demonstrator, data from three different key sources of socio-economic data was selected for inclusion in the ConvertGrid system. Each of these sources provides information on a different set of themes, sampled to different geographical systems.

- 1991 Census Local Base and Small Area Statistics (LBS/SAS) (1991 Census geographies)
- National Statistics Neighbourhood Statistics (1998 ward and local authority district geographies)
- ESRC/JISC 2000 supply of Experian data (2000 postal sector geography)

These datasets were chosen primarily to demonstrate the potential of ConvertGrid to provide integrated access to data sampled on non-contiguous geographies (data from different geographical systems, or from the same system but at different dates, where boundary change has occurred during the intervening period). These datasets were also selected as conversion tables for the different geographies had been generated from the 1999b version of the All Fields Postcode Directory. Unfortunately, as the 1999b version of the AFPD was the only one available when the project started it was not possible to use aggregate statistics from the 2001 Census in the demonstrator.

Three themed use-case scenarios were developed in order to demonstrate potential applications of the ConvertGrid system and to identify which data items from each of the three source datasets should be used in the demonstrator. The themes selected for the use cases were heath, education and crime.

3.3 Development of the ConvertGrid architecture

One of the first tasks was to produce an architecture for the ConvertGrid system that would deliver the required functionality. A number of changes were made to the ConvertGrid architecture specified in the original proposal to reflect a number of issues relating to the development of GSI (Grid Security Infrastructure) enabled OGSA-DAI clients.

- Writing OGSA-DAI clients using the client toolkit is relatively straight forward as long as GSI (Grid Security Infrastructure) is not required. GSI makes writing SOAP clients to an OGSA-DAI Grid service very difficult and requires either using a Proxy

front end to the Grid service or a dedicated API (available in Java and *maybe* Python but ruling out other languages like Perl).

- Dealing with GSI at the Grid service client level means dealing with X509 certification administration which is non-trivial. Also, X509 certification authorization is not required for a demonstrator as Athens could be used to simply control access to the system.
- Querying OGSA-DAI (version 4) enabled data sets can be highly inefficient. A preliminary test shows accessing the Convert Tables using OGSA-DAI by the ConvertGrid business logic is almost ten times slower than by using JDBC. Although this degradation in performance is acceptable at the Web service level - when the actual data conversion is taking place of the final query - it makes writing an efficient dynamic web site problematic.
- Caching query results in OGSA-DAI would probably require overriding the Grid data service classes, hence writing specific code, which could prove to be highly complex.

The revised ConvertGrid architecture detailed in figure 1 represents an attempt to make an equitable compromise between ease of use; implementation complexity; performance and manageability. It proposes to do this by having one web service (running https) that encapsulates the OGSA-DAI complexity. The web service will:

- Provide access to the data sets through simple SQL querying.
- Provide access to the meta-data of the available data sets.
- Offer the actual geographic data convert and fusion operation.
- Enable conversion and fusion of data uploaded by the user.

Web site users or actual client developers will only need an Athens username/password to access and convert data thus eliminating the complexity of handling X509 certificates.

Another cause of change to the ConvertGrid architecture was the decision to use the JISC funded National Grid Service (NGS) (http://www.ngs.ac.uk/) Data Node at the University of Manchester to provide the data Grid infrastructure for the ConvertGrid project. ConvertGrid was one of the early adopters of the NGS service.

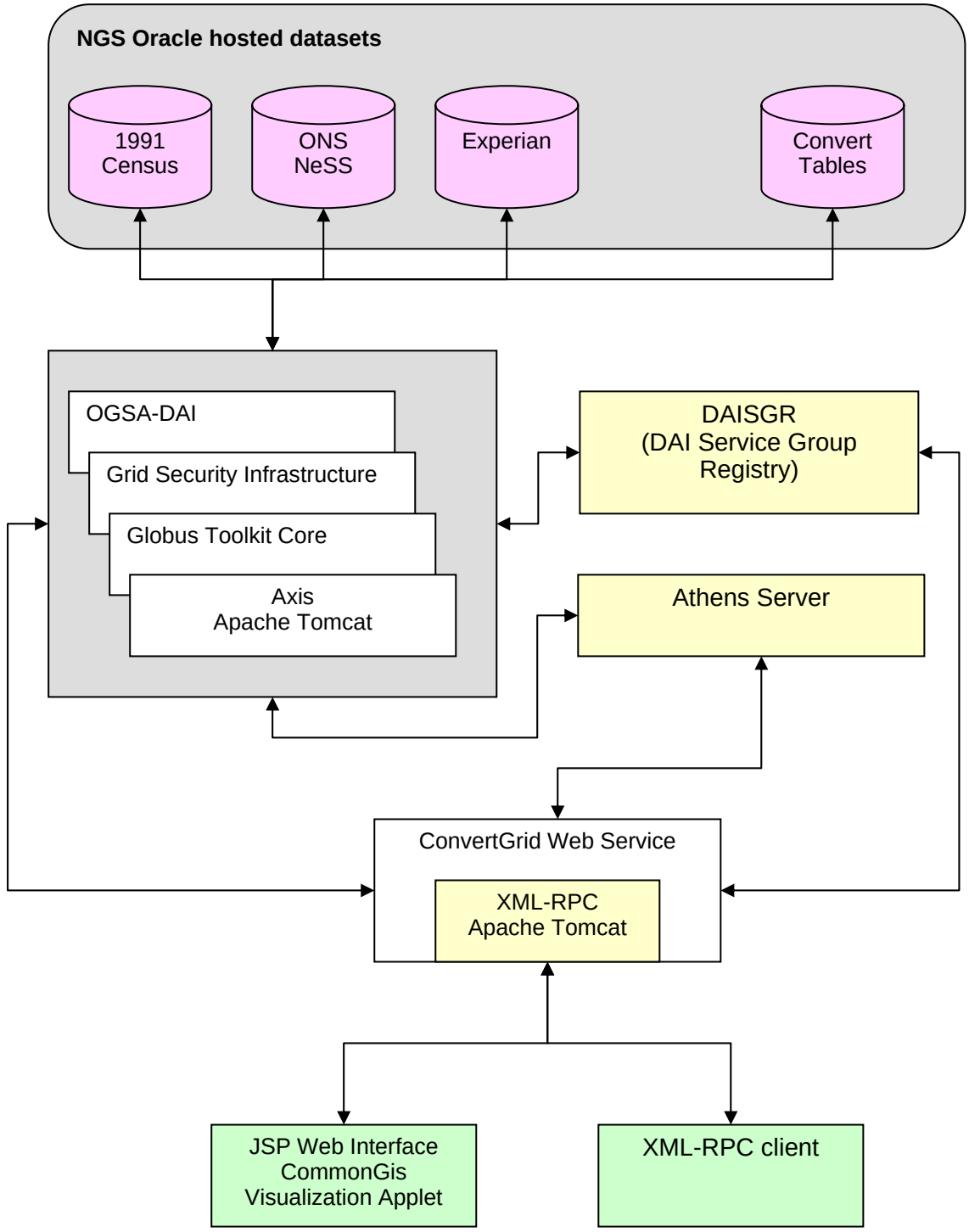

Figure 1. ConvertGrid Architecture

4. Results

4.1 Overview of ConvertGrid functionality

ConvertGrid is a web service which offers the following services:

- Conversion of data sources with different native geographies to a common Target Geography. ConvertGrid can currently convert data in either CSV (comma separated values) or XML (complying with standard Java WebRowSet schema) format. Results, also in CSV or XML, are returned as a string or, as in the case of larger dataset results, streams (FTP/HTTP/GridFTP).

- The ConvertGrid service allows the agglomeration of different aggregate data sources, regardless of their initial geographic format, into a single data source whose geography is determined by the user.

- Several datasets have been Grid-enabled and made accessible for querying through the web service API. These include a subset of the Census 1991 dataset, the Experian Postcode dataset as well as selections from the ONS Neighbourhood Statistics. In addition, Grid-enabled access was provided to the All Fields Postcode Directory (AFPD) derived look-up tables. Consequently, any Grid-enabled dataset, regardless of its location, can easily be added to the ConvertGrid web service simply by notifying the ConvertGrid OGSA-DAI registry. Naturally, the user of the service can also upload data to the service for conversion in combination with the result of ConvertGrid hosted dataset queries or on its own.

- ConvertGrid is accessible through its light client (Web interface) to all registered owners of an Athens account. ConvertGrid is also accessible through its web service API, with examples in various programming languages including Java, Python, PHP and Perl. ConvertGrid and its associated data sets can also be accessed through OGSA-DAI as well as DQP, a middleware component built on top of OGSA-DAI which permits cross querying across federated data sets.

4.2 ConvertGrid Web Service Client

Initially the ConvertGrid Web Service was developed to be accessible through SOAP (Simple Object Access Protocol). Unfortunately, there were problems when calling the Jakarta Axis SOAP library required by the OGSA-DAI client library from the Tomcat web server (which acts as host to the web service). Several days were spent trying to resolve what appeared to be a compatibility issue before raising this with the OGSA-DAI support team. It became clear from the feedback obtained from the OGSA-DAI developers that the problem was more serious than originally thought and wasn't going to be resolved within the time frame for the project.

After several inconclusive trials with various third party Java SOAP libraries the ConvertGrid web service API was made accessible through an XML-RPC library, also running as a Java servlet application inside Tomcat. XML-RPC does not allow for the definition of custom data objects, or classes, which can then be transmitted as is the case with SOAP; however this function was not required by ConvertGrid. XML-RPC has the advantage of being a fairly

simple protocol to implement as it does not use XML name spaces (XMLNS) and only supports standard data types such as string, integers, floats etc.

As a result XML-RPC has proved to be very fast and has the added advantage of being freely available as a programming library for every major language that supports a TCP/IP stack. Several ConvertGrid clients have been developed for demonstration purposes (Java, Python, PHP and Perl) and are downloadable from the ConvertGrid portal site. Also, a thorough (15 pages) hands-on type approach tutorial was written in view of helping developers write their own service clients. Finally, the full ConvertGrid web service API is publicly accessible as is the complete source code documentation of the ConvertGrid project in JavaDoc format.

4.3 ConvertGrid light client (Web front-end)

Following the definition of the API, the development of the web front-end interface was carried out in parallel with the creation of the web service. Initially prototyped in Cold Fusion and SQLServer on a Windows server – the approach used for the Casweb system (http://census.ac.uk/casweb/) used to provide web-based access to the aggregate statistics from the UK censuses of population (Harris, Hayes and Cole 2002) - the front-end was later re-written in JSP (Java Server Pages) and run on the same instance of Tomcat that hosts the web service to provide an Open Source solution.

The web interface offers sufficiently accredited Athens users (see section 4.4) a simple and efficient way of accessing the Grid-enabled datasets. Various drill-down menus offer the user the ability to fine tune the geography which they want their data converted to. In addition to being able to upload user data to the convert service, the interface also allows the user to browse through the Grid-enabled datasets that have notified themselves to the registry (Census 1991, Experian and ONS Neighbourhood Statistics) and select the data to be converted though a series of comprehensive menus (see step by step guide available at https://pascal.mvc.mcc.ac.uk:8443/web/step0Welcome.jsp).

ConvertGrid also offers the possibility to freely-mix data originating from the online Grid-enabled datasets as well as the user upload data thus resulting in a single file characterized by common target geography. If the target geography is a 1991 Census geography, the user may also visualize the result of the conversion with the CommonGIS applet, a thematic mapping tool which runs inside the browser. Figure 2 is a screenshot of CommonGIS cross-classification map showing the relationship between average house price sales (Experian data) and the percentage of 16-19 year olds entering university (Neighbourhood Statistics & Census aggregate statistics) for Greater London. The production of this interactive map is a six stage process within ConvertGrid and took approximately 10 minutes from start to finish. Using this interactive map it is possible to explore interesting spatial patterns, such as those areas with low house prices but high participation rates.

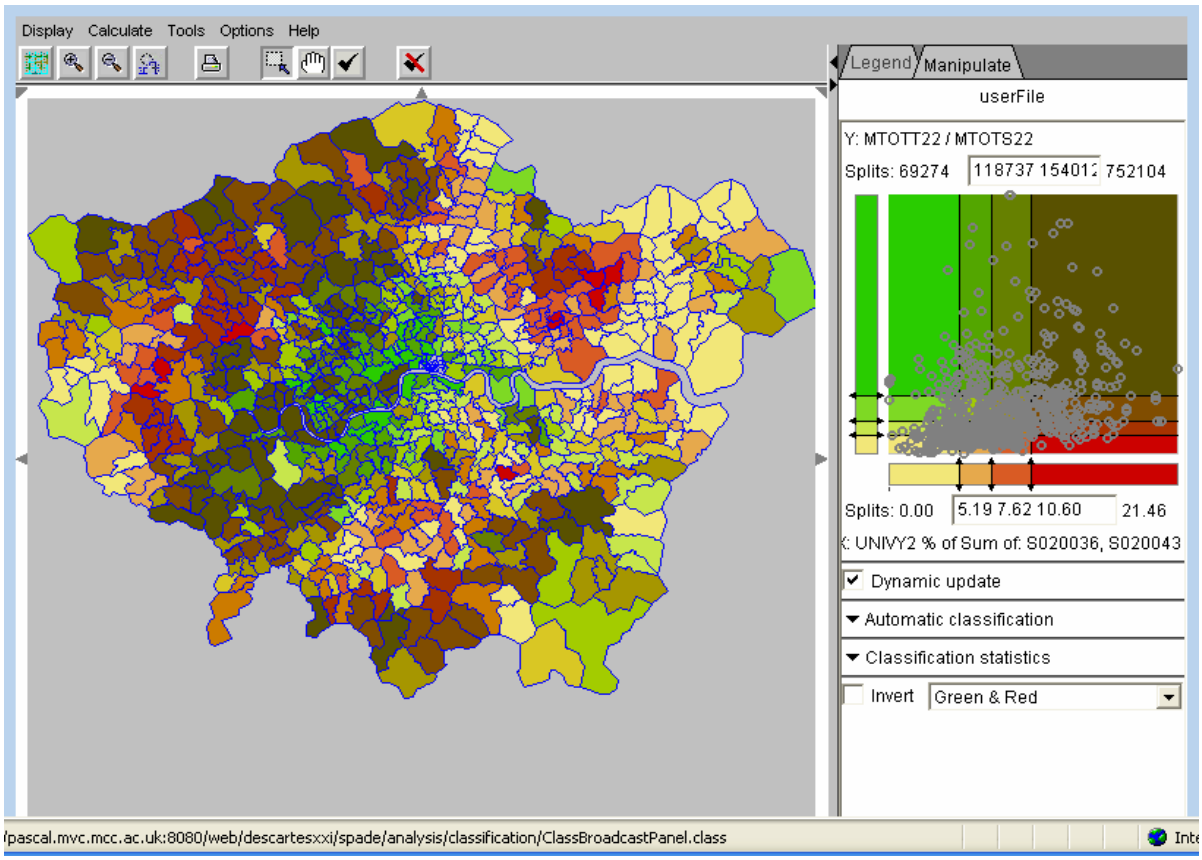

Figure 2. ConvertGrid Data Visualisation Interface

4.4 Security

Due to access restrictions applying to the datasets that have been Grid-enabled and made accessible through ConvertGrid, extra care was put into the design of the security. It was clear from the start that though OGSA-DAI offered a very thorough security mechanism by integrating in its core GSI (Grid Security Infrastructure), the level of expertise required to develop client software making use of a proxy Grid certificate is scarce. Also, adopting GSI throughout would have required that users of the web interface necessarily register and install a UK e-Science certificate which would have been an unjustified overhead for a demonstrator project.

For this reason custom Java code was developed which acts as a proxy to an Athens authentication server. Athens is the system used to provide secure access management to web-based services for the UK education and health sectors (http://www.athens.ac.uk/). This Java code was used by the ConvertGrid service during session creation to authenticate the login/password combination of the user and authorize access to Census data. Further authorization is carried out when accessing the Experian Postal Sector Data as it requires the user to have agreed to special terms and conditions of use.

Once the user is authenticated, a session is created which lasts 30 minutes (counter reset after each user operation). Web services are usually stateless but ConvertGrid has its own session manager, necessary for appending sources files before conversion and fusion. The last query & conversion results as well as the accompanying log file (in XML format) are made available throughout duration of the session.

Owners of an Athens account wishing to use the Web interface (light client) are required to login at the beginning of the session. Message passing from the client to the ConvertGrid Web Service and vice-versa is completely secure as it is carried out through an encrypted connection – HTTP over SSL (Secure Socket Layer) also known as HTTPS. Therefore it is important to stress the fact that no knowledge of security protocols is necessary to develop a client for the ConvertGrid web service and any ordinary XML-RPC library will suffice to allow full use of the service.

Finally, the front-end client to ConvertGrid provides access to most of ConvertGrid's functionality through a series of simple web forms that are served securely by making use of HTTPS.

Application developers who posses a NGS registered UK e-Science certificate can access all Grid-enabled datasets through OGSA-DAI, and this securely through the OGSA-DAI client toolkit and the GSI (Grid Security Infrastructure), by writing fairly simple Java code (see the code registry on the ConvertGrid website for an example). GSI provides a means of identifying accredited users over an encrypted communication channel.

4.5 Grid-enabled datasets

4.5.1 Creating the databases

The three key socio-economic datasets and the AFPD derived conversion tables were Grid-enabled using Oracle and the OGSA-DAI middleware on the NGS (http://www.ngs.ac.uk/). When the project started, there was only OGSA-DAI support for Oracle or MySQL (an open source database) which required data to be migrated from the SQLSever databases used for hosting the various datasets to Oracle on the NGS.

When the development of ConvertGrid began the NGS had not yet installed the Oracle database or OGSA-DAI. A temporary solution was needed for development purposes and so OGSA-DAI and MySQL were installed on a Linux server. Obviously, at this stage only one Grid-enabled dataset was needed for testing OGSA-DAI and the business logic. A subset of the ONS Neighbourhood statistics was transferred from SQL Server to MySQL by means of a comma separated file (generated by the SQL server transfer wizard) which was then read into MySQL after having created the appropriate database schema. A script was written which decompressed the Convert Tables from ZIP format and stored the data in a relational database automatically. The Convert Tables, and the corresponding index information, take approximately 4.5 GB of disk space and were generated in 12 hours.

In parallel to this, the front-end web interface was developed. SQL Server was used to provide the necessary databases and metadata tables. The Census 1991 and the AFPD Convert databases already existed, but the Experian database and the subsets of the Neighbourhood Statistics at ward and district level had to be generated from raw data. Metadata tables had to be created to provide lookups between variable and area codes, and their descriptions. As users were able to choose any areas in the country, for a number of different geographies, some drilldown mechanism was necessary, so area hierarchies had to be created. The relevant ONS SNAC database was utilized to create drilldown hierarchies for the Ward 98 geography. Others, such as Parliamentary Constituencies are accessed via a simple alphabetic index.

When Oracle was finally installed and made available to us through the NGS, it became necessary to transfer all of the datasets from SQL Server and MySQL to Oracle. Installing the

Oracle client on the SQLServer system and then using the the SQLServer wizard to communicate directly to Oracle, seemed to be the most efficient way of doing this. Unfortunately, due to the sheer size of the subset of the 1991 Census dataset it proved to be more problematic than with the smaller datasets. Transferring large amounts of data led to lost connection problems and every table had to be checked manually for missing or duplicate data. Also, due to limitations in Oracle such as the number of columns allowed per table and the size limit of table names and other features such as the handling of mixed case names and minor differences in the SQL commands, major changes had to be carried out before a subset of the Census dataset could be transferred successfully. This underlined the potential amount of work involved in moving from one database system to another.

The use of this amount of Census data used was ambitious compared with the initial plan to use only a small subset of Census data and was significant in highlighting some of the problems that might be encountered in scaling up the system to handle the entire 1991 UK Census. However, to reduce the complexity of the interface development task, only data for England was used for the demonstrator.

Within the convert database there were a number of inconsistencies with the naming of tables and variables (e.g. WARD91 or CENUSWARD91, for 1991 Census wards) which needed resolving. Postcode format problems were also encountered: the formats used in Experian were different to those used in the convert database which required all the Experian postcodes to be reformatted. Also as the convert system can only meaningfully convert counts it was necessary to remove all rates, averages, indices and percentages from the Experian and Neighbourhood Statistics. To expose datasets to the Grid they need to be consistent within themselves and very explicitly defined via their metadata. Standards are vital (e.g. standard postcode formats) to avoid confusion.

4.5.2 Metadata

The ConvertGrid Web Service is tied to an instance of OGSA-DAI, which in turn is tied to an instance of the Grid service registry. The registry contains information on the various Grid services being hosted by the middleware, including Grid data services that encapsulate the datasets, as well as externally hosted Grid services that have notified their existence to the registry.

For ConvertGrid to undertake conversions on a dataset it must parse an XML file which contains information concerning the native geography of the data and, for each table, the column containing the geographic identifier. Fortunately, OGSA-DAI provides a standard way to interface a dataset to an instance of OGSA-DAI and to provide user- defined metadata at the same time. Grid-enabling a dataset to be used by ConvertGrid is simply a matter of writing a short XML configuration file:

```
<?xml version="1.0" encoding="UTF-8"?>
 …
 <!-- Metadata describing the data resource -->
 <metaData>
  <!-- These elements and their contents are optional -->
  <productInfo>
   <productName>Oracle</productName>
   <productVersion>9.2i</productVersion>
   <vendorName>Oracle</vendorName>
  </productInfo>
```

```
<relationalMetaData>
  <databaseSchema
callback="uk.org.ogsadai.dataresource.SimpleJDBCMetaDataExtractor"/>
</relationalMetaData>
<convertGrid>
  <geography>PS99</geography>
  <table name="HOUSEPRICESANNUAL">
    <geoColumnName>AREACODE</geoColumnName>
    <geoColumnNumber>1</geoColumnNumber>
  </table>
  <table name="HHANDPOP">
    <geoColumnName>AREACODE</geoColumnName>
    <geoColumnNumber>1</geoColumnNumber>
  </table>
…
```

The <convertGrid> tag acts as a reference pointer to ConvertGrid when parsing the XML. The <Geography> tag specifies the native geography of the data. Table(s) are defined are defined by their name and the column containing the geographic identifier.

4.6 The ConvertGrid website

The ConvertGrid website (http://pascal.mvc.mcc.ac.uk:9080/convert) currently offers the following documents and services:

- A fifteen page step by step tutorial explaining the creation of a ConvertGrid web service client. The text includes numerous examples in the Python script language.
- A how-to guide on installing and maintaining ConvertGrid and its software dependencies which include Java, Tomcat, Globus, OGSA-DAI, MySQL, Oracle client and the FTP server. Access is currently limited to the ConvertGrid project team members.
- A list of features describing the ConvertGrid service and its web interface.
- A source code repository containing all of the service client examples developed specifically for ConvertGrid. These include clients developed in JSP/Java, Python, PHP and Perl. Also included are the scripts which generate the necessary databases such as the Convert Tables and the National Statistics tables as well as the script which generates the XML configuration file necessary to interface OGSA-DAI to the Census data dataset.
- A documented Web Service API as well as the complete ConvertGrid source code documentation (JavaDoc).
- A web-triggered client which queries the ConvertGrid web service and reports live results. This code is also executed on an hourly basis by a cron daemon and eventual problems reported to the developers by email. Also, a panel in the right column viewable at all times indicates the current status of the FTP, Tomcat and MySQL servers.
- A bug-tracker was also installed and helps the developers of the ConvertGrid project keep track of the problems encountered by themselves and users alike.

6. Future Research Priorities

In order to deploy ConvertGrid as a full web service it was essential that ConvertGrid be adapted to work with the newest releases of Grid software, namely Globus 4. The new Grid technology put forward by the Globus alliance in version 4 of their toolkit is entirely based on SOAP, which is currently the dominant protocol used by web services today. The switch to Globus 4 and the accompanying version of OGSA-DAI would enable ConvertGrid to offer its services through a standard protocol to ALL web service clients as well as the more elaborate Grid clients. Having ConvertGrid adopt SOAP throughout as its main protocol would also mean that the conversion service could be integrated into the new SOAP document model. Finally, a SOAP version of ConvertGrid would be compatible with DQP (Distributed Query Processor based on the OGSA and OGSA-DAI standards.

Based on the experiences of the ConvertGrid project, migrating data to Oracle on the NGS might not be the best approach for Grid-enabling existing social science data services. An alternative approach to Grid-enabling datasets would be to use OGSA-DAI on the NGS to connect to the existing SQLSever databases and non-relational database format datasets maintained by the data provider(s). The advantage of this approach is that it maximizes the investment in the existing data service infrastructure and avoids unnecessary duplication of effort by having to maintain multiple databases systems to support different forms of access.

One of the key features of the census data service at MIMAS is that a lot of the metadata and contextual information required for data selection and extraction is embedded in the current web based interface. One of the issues identified by the ConvertGrid project is that Grid-enabling a subset of the aggregate census statistics has resulted in a separation of the data from the contextual metadata. To resolve this it would be necessary to generate domain specific metadata for the aggregate census statistics which would be held as XML in a Grid service registry. This metadata registry could then be used for resource discovery and also to enable developers to build the appropriate SQL query statements to access specific data items via OGSA-DAI on the NGS. In addition, it would be necessary to develop a tool which would enable the metadata for the extracted data items to be incorporated into the XML data stream returned by OGSA-DAI. This would enable the application developer to be able to process the metadata as well as the data and thus enhance the quality of the information provided to the end user.

7. Conclusion

In conclusion, the ConvertGrid project has demonstrated how Grid technologies can be used to automate very complex workflows and also help to stimulate novel forms of research through promoting increased and more effective use of multiple data sources. However, developing a Grid based service to access, integrate and analyse multiple datasets which is simple to use has required the project team to address a number of key technical and/or methodological challenges. Whilst the use of Grid technologies is currently not well embedded within the social science research community the ConvertGrid project has demonstrated what can be achieved when Grid engineers and data service providers collaborate to overcome barriers to social science research problems. In this context, the project will be working closely with the National Centre for e-Social Science (NCeSS) to ensure that the ConvertGrid project contributes both to capacity building activities within the social science community and the development of an emerging e-Social Science Data Grid.

Acknowledgements

The ConvertGrid project was funded by the Economic and Social Research Council (ESRC) RES-149-25-0008

References

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

Harris, J., Hayes, J, and Cole, K. (2002): 'Disseminating area statistics over the Web', in Rees, P., Martin, D., and Williamson, P. (eds.): *The Census Data System*. John Wiley & Sons, England, pp.113-121

Simpson, L. (2002): 'Geography conversion tables: a framework for conversion of data between geographical units', *International Journal of Population Geography*, **8** (1), pp.69-82. http://www.ccsr.ac.uk/research/documents/ijpgsimpson2002.pdf

Simpson, L. and Yu, A. (2003): 'Public access to conversion of data between geographies, with multiple look up tables derived from a postal directory'. *Computers, Environment and Urban Systems*, 27, pp. 283-307. http://www.ccsr.ac.uk/research/documents/ceussimpsonyu2003.pdf