Data Curation Standards and the Messy World of Social Science Occupational Information Resources

Paul Lambert[1], Larry Tan[1,2], Ken Turner[2], Vernon Gayle[1], Richard Sinnott[3], and Ken Prandy[4]

[1]Applied Social Science, University of Stirling; [2]Computing Science and Mathematics, University of Stirling; [3]National eScience Centre, University of Glasgow; [4]Cardiff School of Social Science, University of Cardiff
paul.lambert@stirling.ac.uk

Abstract. Occupational information resources – data about the characteristics of different occupational positions – play a unique role in social science research. They are of relevance across diverse research disciplines and in numerous disparate contexts. They are also very widely available, typically freely downloadable from academic web-pages. But they are also one of the most uncoordinated types of information resource that social scientists routinely come across. In this paper we describe issues in curating occupational information resources during the GEODE research project (Grid Enabled Occupational Data Environment, http:/www.geode.stir.ac.uk). This project attempts to develop long-term standards for the distribution of occupational information resources, by providing a standardised framework electronic depository for occupational information resources, and by providing a data-indexing service, based on eScience middleware, which collates occupational information resources and makes them readily accessible to non-specialist social scientists.

1 Introduction

1.1 Background

The analysis of occupational positions is a staple component of social science research. In sociology in particular, the connection between occupational and social structures is deeply ingrained in an array of theoretical accounts [33]. Equally, occupational analyses can be found across numerous other research disciplines, as seen for instance in traditions studying labour incomes in economics [28]; the health impacts of occupational inequalities in epidemiology [1]; the evolution of social and occupational associations in social history [31]; and studies of interpersonal relations in social psychology [6].

'Occupational information', as defined here, is used by social scientists to make sense of 'source occupational data'. Occupational information constitutes data about the characteristics of different occupational positions. This may be linked to 'source' data (such as survey questionnaire responses) on the particular occupational positions held by the subjects of analysis. In almost all examples, researchers wish to use occupational information in order to summarise data about their subjects' occupational circumstances in a substantively meaningful but parsimonious way. As a well known illustration, sociologists often wish to use occupational information in order to classify individuals into occupationally based 'social class' groups on the basis of their current occupation. Indeed, deriving occupationally-based social classifications such as social class groupings is a particularly well-developed aspect of occupational information provision. Nevertheless, it is important to appreciate that social scientists make use of many other occupational information resources, another example being statistical databases on occupational circumstances, say on average incomes by occupational groups [21], or the proportion of women in different occupations [14].

1.2 Existing occupational information resources

The common interests shared by many social scientists to summarise source occupational data has not been matched by a shared exploitation of the same methods for processing occupational information. In fact there is little agreement or consistency in social science analyses over which occupational information resources should be favoured from an array of alternatives, or on how precisely to process different occupational information resources in a consistent way. This is the situation which is described as 'messy' in the title of this text.

Thousands of different occupational information resources are available to social scientists. Table 1 describes an illustrative selection, though there are numerous further resources. The abundance can be explained by a number of trends in social science practice. One involves motivations to provide many different types of occupational information, reflecting the wide range of interests in occupational analyses. Other reasons include frequent upgrades and revisions to previously published occupational information resources (typically in response to perceptions of the changing nature of occupational positions over time); as well as growing interest in providing occupational information resources for different countries in internationally comparative analyses. Additionally, increasing access to large-scale data resources which permit the generation of detailed occupational information [13], and increasing willingness to disseminate information resources through internet sites [12], can also explain the expansion of occupational information resources.

Table 1. Selected occupational information resources in the social sciences.

Format	Index units [#]	Output	Documentation	# files / size	ISI citations[1]
1 CAMSIS derivation matrices, www.camsis.stir.ac.uk/ [2001; >10 revisions]					
SPSS / plain text	*OUG; e.s; gender [2000]*	*Scale scores*	√√	*200 / 100kb*	*5*
2.1 ISEI tools, home.fsw.vu.nl/~ganzeboom/pisa [1992; est. 5 revisions]					
SPSS	*OUG [533]*	*Scale scores*	√√	*20 / 50kb*	*61 est.*
2.2 ISEI tools at IDEAS-REPEC, ideas.repec.org/c/boc/bocode/s425802.html [2002; 1 revision]					
Stata	*OUG [533]*	*Scale scores*	√√√	*20 / 50kb*	*31 est.*
3.1 E-SEC matrices, www.iser.essex.ac.uk/esec/ [2005; est. 3 revisions]					
MS-Excel; SPSS	*OUG; e.s. [4000]*	*Social class*	√√	*3 / 100kb*	*0*
3.2 NS-SEC derivation matrices, www.statistics.gov.uk/methods_quality/ns_sec [2001; no revisions]					
MS-Excel; pdf	*OUG; e.s. [3000]*	*Social class*	√√	*6 / 200kb*	*24*
4. Hakim [14] gender segregation codes [1998; no revisions]					
Book	*OUG [400]*	*%female per job*	√	*2 / n.a.*	*38 est.*
5. HISCO occupational labels and codes, historyofwork.iisg.nl [2003; no revisions]					
Book; web html	*OUG [500]*	*OUG content*	√	*1 / n.a.*	*13*
6. O-NET database, http://www.occupationalclassifications-titles.net/ [2001; >10 revisions]					
Web html	*OUG [400]*	*OUG content*	√	*100 / n.a.*	*7*
7. Wright [32] class scheme classification instructions [1985; 1 revision]					
Book	*Job content data*	*Social class*	√	*2*	*100 est.*

- 'Index units [#]': type of index, and average number of distinct index units, in resource.
- '# files / size': approximate number of distinct files, and typical size of each, in resource.
- [] : year of first publication, and number of subsequent updates to contents.
- OUG = occupational unit group; e.s. = 'employment status'.
- Documentation: √ / √√ / √√√ = brief natural language / extended natural language / metadata.

Table 1 conveys some common characteristics of occupational information resources. Important features include that these resources are usually, but not exclusively, distributed through small electronic files from public internet sites; that resource providers usually distribute several distinct electronic files albeit in a coordinated way; that resource providers often update the information they release; and that limited natural language documentation is usually provided by the resource suppliers.

Table 1 also illustrates that there is little coordination among alternative occupational information resources. Perhaps the only common factor is that most resources are oriented around a definition of occupational positions into 'index units' (see 2.1). However, different occupational information resources use different index units. Moreover, different resources tend to supply information in different data formats, and to provide different levels of documentation. Under current practices, researchers wishing to fully evaluate a range of alternative occupational information resources would be required to use multiple software packages and to undertake a variety of data manipulations.

1.3 Current practices and problems in occupational analysis

Social scientists who wish to analyze occupational records are usually faced with two challenges. First, they must navigate and decide among the numerous available occupational information resources. Second, they must implement a connection between the relevant published resource(s), and the occupational records in their own database. Both of these processes have hitherto proved difficult for many social researchers.

One revealing insight into current practices in occupational analysis is the data conveyed in the last column of Table 1, intended to be indicative of the uptake of published occupational information resources[1]. A well disciplined model of social science investigation would see most researchers using published occupational information in a consistent and well documented way [13]. However Table 1 suggests that the practice of occupational analysis is far more 'messy' than this model. Relatively few people, as a proportion of those conducting occupational analysis, appear to utilise these published resources. Indeed, when they do use published resources, Table 1 suggests that users favour those which have more limited documentation and simpler formats (such as the text publications 4 and 7). Several previous authors have noted that, by contrast, social science researchers are much more likely to construct their 'own' occupational information, and deploy it in their own idiosyncratic (and undocumented) style [3], [20], [18].

Another handle on current practices can be gained by reviewing examples of published research which exploits occupational data. A variety of standards can be uncovered. As illustrations, analyses by Bihagen and Ohls [4] and Platt [27] can be heralded as research which maximises the evaluation of candidate occupational information and the provision of documentation. Bihagen and Ohl's work involved linking survey data with three different occupationally based social classifications by using published index files which are described in the text; the analysis incorporated an evaluation of the relative properties of the three classifications considered. Platt's work involved selecting a single occupationally based social classification for analysis, on the basis of a sequence of explicit decisions about the quality of its documentation and comparability; this decision making process was discussed in detail in an online appendix. However it should be recognised that the effort and skills involved in these implementations are substantial, and relatively few social scientists have demonstrated the diligence illustrated by these studies.

The analyses of Archer and Francis [2], Dixon and Paxton [8] and Modood [22] may be presented as more problematic examples of occupational research. All implement classifications of occupational positions on the basis of occupational information. However Archer and Francis's shortcoming involves their designation of an occupational class classification based upon their own judgment, which generates a scheme which is not replicable and may not be readily compared with other published analyses. Dixon and Paxton's weakness involves their attempt to synthesise results from a series of occupationally based social class classifications which are not equivalent, without providing details on the implications of alternative schemes or their comparability. Modood's text similarly lacks details on the specification of a simplistic occupationally based social class classification which is used to help explain patterns of educational attainment (p. 300ff). The problems are exacerbated in this instance because, as Modood notes, complexities to the labour market situations of the groups studied are ignored by the classification used. Implicitly, there is no possibility for a reader to anticipate the magnitude of the 'true' effects, which may have been revealed had more sophisticated occupational information been employed. In each case, the limitations associated with these outputs suggest that the researchers were not in a position to comfortably review a wider range of potentially relevant occupational information resources, nor undertake and document a clearly defined linking process.

1.4 Strategies for managing occupational data.

The difficulties confronted by social researchers studying occupations may well result from the large volume of alternative occupational information resources and the lack of coordination between them. One reaction could be to impose standardization on the collection and analysis of occupational data, for instance by enforcing data collectors to code records into a standardized occupational scheme, and by asserting that certain occupational information resources should be preferred over rivals. This strategy is well illustrated by 'universal' approaches to occupationally based social classifications, in which it is

[1] The number of ISI indexed journal articles citing the relevant occupational information resource's documentation in their bibliography (calculated from Web of Knowledge citation statistics, www.wok.mimas.ac.uk). The figures for the ISEI tools show an estimated fraction of 92 studies citing the ISEI documentation. The figures for the Hakim and Wright texts are estimates derived from the total number of citations of the books.

asserted that occupational structures across different countries and time periods are broadly stable, and that a single occupationally based social classification is adequate for all research investigations [15: 2]. However, a universal approach to occupational information resources can be demonstrated to be both theoretically and empirically unsatisfactory, since it is prone to neglect patterns of occupational change [19]. Moreover, the pluralistic theoretical traditions of social scientists [33], along with inconsistencies in the practices of existing social researchers, suggest that a universal approach to occupational information is, in practice, unattainable[2].

Instead a pluralistic approach to managing diverse occupational information resources is more attractive. Here, researchers' access to alternative resources may be facilitated and encouraged, and standards of explicit documentation and evaluation fostered. One option could be to provide informative textual comparisons of occupational information resources. Several texts have provided focussed reviews of selected resources [14: Annex A]. However, there has been no widely accepted systematic summary of all occupational information resources, and prospects for such an undertaking would seem unlikely, given the variety and volume of resources involved.

An alternative is a computer-based facility for describing occupational information. The GEODE project (www.geode.stir.ac.uk) seeks to provide an online database which collates data on occupational information resources and distributes it across the social science research community. In particular, the GEODE project exploits the capabilities conferred by eScience computing, notably security, data abstraction and virtual organizations. It attempts to develop long-term standards for the distribution of occupational information resources, by providing a standardised framework-based digital depository for occupational information resources, and by providing a data-indexing service, based on eScience middleware, which collates occupational information resources and makes them readily accessible to non-specialist social scientists.

2. Curation of Occupational Information Resources

2.1 Metadata requirements

The existing arrangements for the distribution of occupational information resources exhibit a clear shortcoming, namely the absence of consistently structured metadata. It is well recognised that consistent standards of data curation through metadata enable rapid navigation and processing of information resources (http://www.dcc.ac.uk/resource/curation-manual/). This is particularly true in the context of Grid enabled datasets [7]. Therefore a first objective of the GEODE project has been to establish a framework for the curation of occupational data.

Following earlier recommendations on e-Social Science standards [7], and in line with prevailing practices in curation of other social science datasets [5], GEODE uses a data curation structure based upon the Michigan Data Documentation Initiative (DDI, www.icpsr.umich.edu/DDI/). This standard is attractive because the storage of metadata in a DDI format allows ready integration with the data manipulation processes also catered for in GEODE (see section 3). The DDI offers a generic set of XML tags which can be used to curate in a consistent manner a large range of social science data. The GEODE project is concerned with a limited range of metadata statements, those required to adequately curate the small data files typical of occupational information resources. Moreover, as such data files are often updated over time, there is motivation to find DDI-based standards of curation that are relatively quick to implement.

GEODE concentrates upon a prescribed subset of DDI tags, referred to as the 'GEODE-M' metadata standard. A review of existing occupational information resources was undertaken in order to establish which information was most important to generating metadata on the occupational records. Three structural contexts were established:

Index schemes for source occupational data.
In most social surveys, a textual description of the occupational title and circumstances is taken as the initial source occupational record. This information may be stored as free text. However, more commonly it is translated into an index of occupational positions, usually a location within an

[2] Indeed, almost forty years ago, Bechhofer's review of the use of occupational information in sociology bemoaned the abundance of, and inconsistencies between, occupationally based social classifications, noting that *"..researchers are advised not to add to the already existing plethora of classifications without very good reason"* [3: 118]. However since that recommendation, the number of new classifications has increased steadily.

'occupational unit group' (OUG) scheme. In most countries, prescriptive documents are available which show how occupational descriptions may be assigned to numerically standardised occupational schemes, such as OUG systems [24], [17], or industrial sector classifications [23]. In several cases, computer software is available to allow rapid classification of textual occupational descriptions into numerical OUG locations[3]. Due to the well-developed nature of this aspect of occupational information handling, the GEODE strategy is to assume that all occupational records have been located within some form of published occupational index scheme.

Three types of source occupational data are most commonly recorded. One concerns the classification of occupational titles into an OUG scheme; it is this type of occupational data which has the widest range of occupational information resources associated with it. Another concerns the industrial sector location of the occupation. As indicated above, standardized index schemes are widely used for classifying occupational titles and industrial sectors. A third type of data is most usually described as 'employment status', and concerns the ownership of the occupational site and circumstances of the employment contract. A variety of national and international employment status indexes exist [9], although many statistical agencies use bespoke employment status questions. In addition to these more common types of record, many studies also hold additional data on the occupational position held by an individual – examples include the normal time and days of work; as well as aspects of the work process such as the extent of supervision experienced.

In seeking to provide facilities for the curation of occupational information resources and their relation to source occupational data, the GEODE project takes as its starting point the assumption that source occupational information has been recorded in the format of a published occupational index scheme such as an occupational unit group (OUG) system (see also 1.2 above). This proves to be an important assumption since published occupational index schemes exhibit the idealised features of a 'standard category' <stdCatgry> record within the DDI system. The declaration of occupational index schemes as standard categories means that connections between occupational information resources, and source occupational data, can in principle be fully leveraged simply by searching for matching combinations of the relevant index scheme(s).

The declaration of a DDI 'standard category' requires reference to further details on each index scheme. In principle this would allow any categorization to serve as an index system. However, the uneven evolution of occupational information resources in the social sciences means that there are several published conflicts between the precise definitions of index schemes. A comprehensive listing of occupational index units would be beneficial in order to allow immediate specifications on the scope of a given standard category. Within GEODE, this is achieved through the manual publication of a listing of occupational index measures at www.geode.stir.ac.uk/ougs.html.

Context of occupational data.

Occupational information resources refer to a wide array of different 'contexts'. Most frequently, resources are associated with contexts defined by different nations and/or different time points. For instance, OUG index classifications and translations are available for different contemporary countries [24 cf. 30] or in cross-nationally comparative contexts [17], and they are published within and between countries with relevance to different time periods [24], [26], [25]; [17], [16]. However, other social contexts may also be used to delimit the coverage of occupational information resources – some resources apply only to the occupations of male or female respondents respectively, or only to other particular social groups[4]. Within the DDI scheme, several tags may be used to define the appropriate context of a given occupational information resource, all of which can be suitably located within the 'study information' <stdyInfo> section of the metadata.

Reference unit of occupational analysis.

A third issue in the recording and processing of source occupational data concerns the 'unit of analysis' to which the occupational information is to be applied. A well known debate within sociological literature concerns whether the occupational class of an individual is best understood in terms of their own current occupation (if working), or by incorporating information on previous occupations or the occupations of household sharers such as a spouse. Although there have been some recommended (but contested) principles for summarizing occupational data [11], the permutations associated with the 'reference unit' for occupational data rapidly become very complex (such as how to adequately describe a career sequence of occupational positions; or how to merge occupational records from multiple household sharers). The data management tasks involved in such data complexities are

[3] E.g. Computer Assisted Structured Coding Tool, www2.warwick.ac.uk/fac/soc/ier/publications/software/cascot/.
[4] For instance the HESA scheme for graduate level occupations, www.hesa.ac.uk/manuals/05018/05018a04.htm.

substantial and arguably have prevented many researchers from adequately exploiting their source occupational data. For instance, it is argued that most sociological researchers use the more easily implemented 'individual' occupational measure, despite overwhelming empirical support for incorporating household level records [18]. Allowing for potentially different reference units is highly attractive; as noted below, the DDI curation of 'variable groups' readily allows data-matching programs to assign multiple linkages.

2.2 GEODE-M metadata standard

The GEODE-M customized metadata standard incorporates entries in each of the five component structures of the DDI. These cover a production statement for the metadata itself; statements on the generation of the occupational information resource; statements describing the data file(s) associated with the resource; data describing the content of the data file(s) of the resource; and space for optional additional statements.

Segments from GEODE-M – example of key DDI XML-tags

```
 <codebook>
<docDscr> … <distStmt> <contact email="pl3@stir.ac.uk"> Paul
    Lambert</contact> </distStmt>
    <prodDate date="2006-07-19" >July 19, 2006</prodDate>
                        …  </docDscr>
<stdyDscr> … <titl>CAMSIS scales for the UK using SOC2000</titl>
    <IDNo agency="GEODE">131</IDNo>
    <distrbtr URI="http://www.camsis.stir.ac.uk">Cambridge Social
    Interaction and Stratification Scales website</distrbtr>
     <stdyInfo> <!-- information about the data context -->
     <sumDscr> <timePrd  event="start" >2000</timePrd>
     <nation abbr="GB">United Kingdom</nation>  </sumDscr>
     </stdyInfo>        … </stdyDscr>
<fileDscr id="gb91soc2000.sav"> …
    <fileName id="gb91soc2000.sav">gb91soc2000.sav</fileName>
                        … </fileDscr>
<dataDscr> …
    <varGrp name="indexs" var="soc2000s ukempsts stdempsts" >
    <concept>Index term</concept> ... </varGrp>
    <varGrp name="outcomes" var="MCAMSISs FCAMSISs">
    <concept>Occupational information</concept> </varGrp>
    <var ID="soc2000s" file="gb91soc2000.sav" >
    <stdCatgry uri="http://www.geode.stir.ac.uk/ougs.html#soc2000">
    Standard Occupational Classification 2000</stdCatgry></var>
                        … </dataDscr>
<otherMat> …            … </otherMat>
 </codebook>
```

The GEODE-M standard is devised in such a way as to minimize the requirements for describing occupational information resources, whilst successfully drawing out the salient identifying features of those occupational information resources which have been reviewed. Figure 1 illustrates the essential contents of a GEODE-M entry, in this example describing a data resource available from the CAMSIS project webpages (www.camsis.stir.ac.uk). The figure shows that only a handful of information records need be assigned to curate an occupational information resource. These cover a contact name for the supply of metadata; a title statement for the resource itself and data on the location and date of publication of the resource; a specification identifying the file or files being curated; and statements identifying the variables contained within the file, making the crucial allocation of variables into appropriate 'variable groups'.

A critical feature of the DDI standard is the specification of 'variable groups'. These define the nature of the occupational information. They identify all information as either an 'index' measure or an output measure (the same variables could be included in both groups). This separation allows rapid indexing of the occupational information resources according to which index variables are included. By exploiting the 'standard category' <stdCatgry> statement, index variables may be coordinated within a single system of occupational resources.

An appealing feature of the DDI format is the repeatable nature of the file description and data description elements. The former allows occupational information resources which supply data simultaneously in more than one data file to be curated as a single body, and resources searched across the range of files. The latter allows multiple variable definitions within each data file to be specified, with comparable content between variables from different files documented if necessary.

The GEODE-M metadata standard serves to indicate aspects of occupational information resources which allow standard index searching and linkage exercises (see section 3). It is intended to allow rapid curation of many occupational information resources, since many potential data suppliers from social science backgrounds are unlikely to be sufficiently motivated to spend long periods curating their own data for other user's benefits. The GEODE-M specification requires as a bare minimum only eleven information statements (underlined in Figure 1; other statements are generated automatically). This allows resources to be deposited to the GEODE index service with minimal manual curation. Files may be submitted to the site through an entry portal which features a short Java proforma collecting the minimum GEODE-M statements.

Nevertheless, additional metadata will improve the quality of data curation and accessibility for index searching and linkages. The GEODE strategy allows further curation of metadata for any deposited resources by both the original depositor and members of the GEODE project. Further details may be voluntarily contributed (through editing of the XML record) to extend the curation process.

3. Conclusions

The innovation associated with the GEODE-M DDI curation concerns its usage to interlink occupational information resources in a Grid enabled data environment. GEODE uses the OGSA-DAI middleware (Database Access and Integration, www.ogsa-dai.org) to provide a framework for these services [29]. The system enables curated occupational resources to be connected and exposed to a virtual organization providing data indexing and matching services. In GEODE, these services may be accessed by non-specialist users through the design of a user-friendly 'portal' interface to the databases which uses GridSphere (www.gridsphere.org).

It is these services which offer substantial improvements in the handling of occupational information amongst social science users. These improvements occur in broadly two contexts:

Robust reviews of occupational index records and documentation

The previous provision of occupational information resources has required users to search diverse resources for data stored in exactly equivalent index units to those on the user's source occupational data. As we have indicated above, there are numerous occupational index schemes in currency, because published occupational index schemes are regularly revised and updated over time; because alternative schemes are available for alternative contexts such as different countries and time periods; and because occupational index schemes are usually designed in such a way as to incorporate alternative levels of detail on the occupational location[5]. Moreover, the numeric format used for recording locations within occupational index schemes is sometimes inconsistent. In some schemes trailing zeros are used to indicate subgroup membership instead of hierarchical truncation (for example SOC 2000 'major group 5' may also be indicated as '5000'). Equally, some schemes are recorded in text formats and/or with decimal markers in order to distinguish truncation in occupational data. For example, the ISCO-88 codes generated by the CASOC software [10] are generated in 'string' format, whilst the UK 1980 classification [25] is commonly recorded as a 5-digit classification featuring decimals between the third and fourth digit. The practical upshot of the range of alternative occupational index schemes and formats is that most occupational information resources are readily available for a very limited range of index schemes. Researchers have previously been pushed into selecting occupational information on the basis of an exact match in index variable formats. However the GEODE use of standard category statements opens up possibilities to declare relations between different index variables, bridging the gaps generated by formatting inconsistencies.

Rapid implementation of secure file-matching procedures

The substantial impediment associated with previous applications in occupational research concerned user difficulties in implementing the software specific linkage between source occupational data and a published occupational information resource. Through its exploitation of OGSA-DAI systems, the GEODE service offers a framework for conducting this linkage in an automated but secure way (recognizing that source data is usually highly sensitive). The mechanics of this linkage hinge on

[5] For example, in the UK SOC-2000 classification [24], an occupation may be noted as 'unit-group' '5232 Vehicle body builders and repairers', but it could alternatively be recorded as 'minor group' '523 Skilled metal and electrical trades: Vehicle trades', or as the 'major group' '5 Skilled trades occupations'.

identifying the index linking variables available in the source data, a process enabled by the specification of standard categories for occupational index variables.

Acknowledgments
This research is supported by an ESRC 'Small Grant in e-Social Science', RES-149-25-1015.

References

1. Arber, S. Comparing inequalities in women's and men's health: Britain in the 1990s, *Social Science and Medicine, 44(6), 773-787,* 1997.
2. Archer, L. and Francis, B. Challenging Classes? Exploring the role of social class within the identities and achievement of British Chinese pupils, *Sociology 40(1), 29-49,* 2006.
3. Bechhofer, F. Occupations, in: Stacey, M. (ed) *Comparability in Social Research,* 94-122, Heinemann, 1969.
4. Bihagen, E. and Ohls, M. The glass ceiling - where is it? *Sociological Review,* 54(1), 20-47, 2006.
5. Blank, G., and Rasmussen, K.B. The Data Documentation Initiative: The Value and Significance of a Worldwide Standard. *Social Science Computer Review 22(3), 307-318*, 2004.
6. Burchell, B., Day, D., et al. *Job Insecurity and Work Intensification: Flexibility and the Changing Boundaries of Work,* York Publishing (Joseph Rowntree Foundation), 1999.
7. Cole, K., Schurer, K., Beedham, H., Hewitt, T. *Grid Enabling Quantitative Social Science Datasets – A Scoping Study,* ESRC/JISC, 2003.
8. Dixon, M. and Paxton, W. The State of the Nation: An audit of social injustice in the UK. In: Pearce, N. and Paxton, W. (eds) *Social Justice: Building a Fairer Britain,* 21-61, Politicos, 2005.
9. Elias, P. Status in employment: A world survey of practices and problems, *Bulletin of Labour Statistics 2000-1, xi-xix*, 2000.
10. Elias, P., Halstead, K., Prandy, K. *Computer Assisted Standard Occupational Classification*, HMSO, 1993.
11. Erikson, R. Social class of men, women and families, *Sociology, 18, 500-514,* 1984.
12. Ganzeboom, H.B.G. *Tools for Deriving Status Measures from ISKO-88 and ISCO-68. http://home.fsw.vu.nl/~ganzeboom/PISA/, [accessed 1.10.06]*, Amsterdam University, 2006.
13. Goldthorpe, J.H. Progress in sociology: The case of social mobility research. In: Svallfors, S. (ed) *Analyzing Inequality,* Stanford University Press, 2005.
14. Hakim, C. *Social Change and Innovation in the Labour Market*, Oxford University Press, 1998.
15. Hout, M. and DiPrete, T.A. What we have learned: RC28s contributions to knowledge about social stratification. *Research into Social Stratification and Mobility, 24(1), 1-20,* 2006.
16. International Labour Office (ILO), *International Standard Classification of Occupations, Revised Edition 1968*, International Labour Office, 1969.
17. International Labour Office (ILO), *ISCO-88: International Standard Classification of Occupations*, International Labour Office, 1990.
18. Lambert, P.S. Handling occupational information. *Building Research Capacity, 4, 9-12,* 2002.
19. Lambert, P.S., Prandy, K., and Bergman, M.M. Specificity and universality in occupation-based social classifications. Paper to: *European Association for Survey Research, Barcelona, 18-22 July*, 2005.
20. Marsh, C. Occupationally based measures. In: Jacoby, A. *The Measurement of Social Class*, pp. 1-47, Social Research Association, 1986.
21. McKnight, A. and Elias, P. A database of information on unit groups of the Standard Occupational Classification. In: Rose, D. and O'Reilly, K., (eds.) *Constructing Classes,* pp. 116-145, ESRC/ONS, 1997.
22. Modood, T. The educational attainments of ethnic minorities in Britain. In: Loury, G.C. et al (eds) *Ethnicity, Social Mobility and Public Policy, pp. 288-308,* Cambridge University Press, 2005.
23. Office for National Statistics (ONS), *UK Standard Industrial Classification of Economic Activities 2003,* The Stationary Office, 2003.
24. Office for National Statistics (ONS), *Standard Occupational Classification 2000: Volume 1. Structure and descriptions of unit groups,* The Stationary Office, 2000.
25. Office for Population Census's and Surveys (OPCS), *Classification of Occupations 1980,* HMSO, 1980.
26. Office for Population Census's and Surveys (OPCS), *Standard Occupational Classificaton,* HMSO, 1990.
27. Platt, L. *Migration and Social Mobility: The Life Chances of Britain's Minority Ethnic Communities,* The Policy Press, 2005.
28. Routh, G. *Occupation and Pay in Great Britain, 1906-79,* MacMillan Press, 1980.
29. Tan, K.L.L., Gayle, V., Lambert, P.S., Sinnott, R.O., Turner, K., GEODE – Sharing occupational data through the Grid, paper to: *5th UK eScience All-Hands Meeting, Nottingham, 18-22 September,* 2006.
30. US Department of Labor, *Standard Occupational Classification Manual (2000 ed),* Washington DC, 2000.
31. van Leeuwen, M.H.D., Maas, I., and Miles, A. (eds) *Marriage Choices and Class Boundaries: Social Endogamy in History,* Cambridge University Press, 2005.
32. Wright, E.O., *Classes,* Verso, 1985.
33. Wright, E.O. (ed) *Approaches to Class Analysis,* Cambridge University Press, 2005.

Author/s:
Lambert, Paul; Tan, Larry; Turner, Ken; Gayle, Vernon; SINNOTT, RICHARD; Prandy, Ken

Title:
Data curation standards and the messy world of social science occupational information resources

Date:
2006

Citation:
Lambert, P., Tan, L., Turner, K., Gayle, V., Sinnott, R., & Prandy, K. (2006). Data curation standards and the messy world of social science occupational information resources. In 2nd International Digital Curation Conference, Glasgow, UK.

Publication Status:
Published

Persistent Link:
http://hdl.handle.net/11343/28841

File Description:
Data curation standards and the messy world of social science occupational information resources