

# Researcher-Led Development of E-Research in the Social Sciences: The Case of an E-Social Science Pilot Demonstrator Project

by Bridgette Wessels and Max Craglia  
University of Sheffield; Institute for Environment and Sustainability

*Sociological Research Online*, 15 (2) 7  
<<http://www.socresonline.org.uk/15/2/7.html>>  
10.5153/sro.2095

Received: 5 Oct 2009 Accepted: 3 Mar 2010 Published: 30 Apr 2010

---

## Abstract

The introduction and use of information and communication technologies (ICT) in the process of research is extending beyond research management into research practice itself. This extension of the use of ICT in research is being termed as e-research. The characteristics of e-research are seen as the combination of three interrelated strands, which are: the increased computerization of the research process; research organized more predominantly in the form of distributed networks of researchers, and a strong emphasis on visualization. E-research has become established in the natural sciences but the development of e-research in relation to social sciences is variable and less pervasive. The richness of the social sciences and their variety of practices and engagement in diverse fields of study mean that e-research as utilized in the natural sciences cannot be easily migrated into the social sciences. This paper explores the development of e-research for the social sciences. The paper is based on an ESRC funded e-social science demonstrator project in which social scientists sought to shape the use of Grid ICT technologies in the research process. The project is called: 'Collaborative Analysis of Offenders' Personal and Area-based Social Exclusion': it addresses social exclusion in relation to how individual and neighbourhood effects account for geographical variations of crime patterns and explores the opportunities and challenges offered by e-research to address the research problem. The paper suggests that if e-research is driven from the needs of social research then it can enhance the practice of social science.

---

**Keywords:** *E-Research; Social and Area-Based Exclusion; Young People at Risk of Crime; Information and Communication Technologies; Collaborative Research; Interdisciplinary Research*

---

## Introduction

**1.1** The introduction and use of information and communication technologies (ICT) in the process of research is extending beyond research management into the practice of research. The extension of the use of ICT in research is being termed 'e-research'. E-research has become established in the natural sciences but the development of e-research in relation to social sciences is variable and less pervasive. The richness of the social sciences and their variety of practices and engagement in diverse fields of study mean that e-research as utilized in the natural sciences cannot be easily migrated into the social sciences. Considerable research funding has gone into exploring and developing e-research for the social sciences. One example from the UK is the Economic and Social Science Research Council (ESRC) 2004-2009 national funding programme for e-social science (c.f. Woolgar, 2003).

**1.2** In this paper we explore how e-research can be adapted for research in social science. A key dimension in the process of developing ICT within research methodology is the roles of social scientists in shaping e-research and informing the design of ICT-based tools. An ESRC-funded e-social science demonstrator project called 'Collaborative Analysis of Offenders' Personal and Area-based Social Exclusion'<sup>[1]</sup> is an example of the work of social scientists in shaping ICT in the research process. The project had two main objectives: to consider the theoretical and policy implications of social exclusion in relation to how individual and neighbourhood effects account for geographical variations of crime patterns; and to explore the opportunities and challenges offered by the Grid from a socio-technical perspective, i.e. how different disciplines and theoretical traditions could engage with this new way of working, and shape emerging ICT. This paper focuses on the second objective, addressing the process of developing e-research by drawing on the methodological and analytical work of the researchers in the project (for full details of the analytical and modelling components see Craglia et al., 2005). The structure of the paper is: first, defining

e-research; second, the practical development of an e-social science demonstrator; and third, a discussion of the findings and their significance in evaluating e-research followed by the conclusion.

## **Defining e-research**

**2.1** Some commentators are pointing out changes in research process and practice that may result in transformations in scholarship (Jankowski, 2009). The broad character of these changes emerge from the combination of three interrelated strands: the increased computerization of the research process; research becoming organized more predominantly in the form of distributed networks of researchers, and a strong emphasis on visualization. These changes to research are termed 'e-research' (Jankowski, 2009, p. 7). There are precursors to this term, for example Nentwich (2003) uses the term 'cyberscience' (originally developed by Wouters in 1996) to situate research practice in the age of the Internet. He argues that cyberscience is: 'scholarly and scientific research activities in the virtual space generated by the networked computers and by advanced information and communication technologies' (Nentwich, 2003, p. 22). This conceptualization includes the practices of the research process as well as ICT-based tools.

**2.2** This term was replaced with 'e-science' in the European context and 'cyberinfrastructure' in the US context. In 1999, the Director of the UK Office of Science and Technology, John Taylor, used the term 'e-science' to support a large-scale funding programme to develop Grid computing and large-scale data processing techniques for the natural sciences. E-science in this context is defined as 'global collaboration in key areas of science and the next generation of infrastructure that will enable it' (John Taylor, Director General of the Research Councils, UK).<sup>[2]</sup> More specifically, e-science involves the coordination of geographically dispersed computing and data sources based on Grid infrastructure. The Grid refers to emerging hardware architecture, software, and standards that are being developed to support the sharing of resources for collaborative working between groups of scientists. The Grid facilitates the combination of extensive connectivity with massive computer power and vast quantities of digital data to support research, and is mainly developed within the natural sciences.

**2.3** In 2000, the ESRC launched a programme to contribute to the e-science initiative and to explore the use of the Grid in the social sciences. This involved funding eleven demonstrator projects and a National Centre for e-Social Science (NceSS) linked with a series of nodes (Halfpenny et al, 2009, p. 74). In the first instance e-social science was defined as the 'collaboration between computer scientists and social scientists to design and develop middleware in order to address social scientists' substantive research problems in new ways that recognize more fully the complexity of economic and social activities' (NceSS, e-Social Science newsletter Issue 1, Summer 2005).

**2.4** As stated in the introduction, a key issue in this development relates to how adaptations of Grid technologies can interact meaningfully within the diversity of social science research. It is in this context that Beaulieu and Wouters' (2009) development of the term e-research is useful because it focuses more broadly on research which uses new media and electronic networks for a range of research practices within the social sciences and humanities. This focus moves developments in, and analysis of, changing research practice beyond the main focus in e-science (which concentrates on large-scale data processing across distributed research networks) to include the variety and richness of research practice in the social sciences.

**2.5** It is necessary to consider some empirical case studies to understand how e-research can be developed for the social sciences. To gain insights into the development of e-research, we consider one of the eleven ESRC-funded e-social science pilot demonstrators in the UK. The 'Collaborative Analysis of Offenders' Personal and Area-based Social Exclusion' project specifically focused on the roles of social scientists in shaping e-research and it did this in an applied setting of a social science project. The strength of this approach is that it does not abstract debate about changes in methodology in e-research out of research practice but instead explores changes through the work of researchers in specific projects.

## **The practical development of an e-social science demonstrator**

**3.1** Given that the ethos of the pilot demonstrator was to explore the practice of developing e-research, the substantive social science focus of the project is important because it frames the work done in the project. The central driver of the 'Collaborative Analysis of Offenders' Personal and Area-based Social Exclusion' was to assess to what extent individual and neighbourhood effects account for the geographical variation of crime patterns. This issue has been at the core of environmental criminology since the early work of the Chicago School of Sociology in the 1920s, and the theoretical debate on the relative importance of individual, family, school, social ties, and neighbourhood factors to crime patterns continues to the present day (Wilson, 1978; Friedrichs & Balsius, 2003; Sampson, Raudenbush and Shapiro, 1997). Given that evidence regarding the influence of area effects remains unclear and difficult to quantify, the Grid provided an opportunity to undertake large-scale data analysis over distributed information resources. This promised to

provide new insights into this key research issue, which has both theoretical and policy implications. The multi-dimensionality of the research problem suggested the need to take an interdisciplinary approach which, in principle, collaborative e-research has the potential to facilitate. With these considerations in mind, the project had two main objectives:

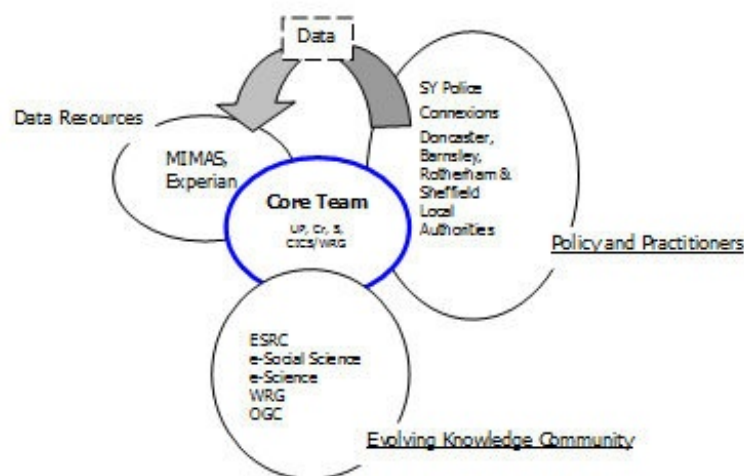
1. To explore, quantify and model the spatial distribution of crime in relation to socio-economic and neighbourhood characteristics based on user-driven applications of Grid technology.
2. To reflect critically on the evolving relationship among social scientists, technologists, and the Grid as input to the development of training material and the further deployment of Grid technology in the social sciences (Wessels and Craglia, 2007).

**3.2** Three main groups of actors took part in the project: (1) the core team of academic researchers, (2) the project partners and data providers from the regional policy-making community, and (3) the support services of external private sector data-suppliers, university Grid infrastructure and tools, and Web services developers. The team came together through pre-proposal meetings to identify the expertise required for the research and development, and the project partners were selected according to the data and expertise they could provide. The authors of this paper were part of the core academic team: Wessels was the ethnographer in the project, and Craglia was the Principal Investigator (PI).

**3.3** The core research team was composed of three urban planners with Geographic Information Systems (GIS) expertise (one of whom was the PI), two criminologists, two sociologists (including the ethnographer), and one computer scientist (Grid Officer) with special responsibility for Grid computing applications. The project partners and data suppliers included South Yorkshire Police, who provided a unique combination of highly confidential data on the location of known offenders, offences, and victims over a five-year period (1998-2003). The four Local Authorities in the region (Barnsley, Doncaster, Sheffield and Rotherham) provided information on their policies that address 'social exclusion' and 'anti-social' behaviour, including policy boundaries. The Government agency, South Yorkshire Connexions, provided data on young people not in employment, education, or training.

**3.4** The support services group involved Experian (a private company that produces the Mosaic (UK) geo-demographic classification), which made its data available at no cost to the project. The University of Sheffield provided support and tools to access the White Rose Grid (WRG) computing infrastructure (which involves collaboration with the Universities of York and Leeds), and the Open Geospatial Consortium (OGC) supported the development of a platform based on its Web service specifications for sharing the results of the project among the stakeholders. Other data sources included the Census, accessed through the University of Manchester academic services (MIMAS), and the Index of Multiple Deprivation 2004 (published by the UK Communities and Local Government Department).

**3.5** The development of e-research in this context involved the collaboration of four sets of actors and resources: the core team, data and data resources, policy-makers and practitioners, and the evolving knowledge community (see Figure 1.):



**Figure 1.** Collaboration and Actors in the Pilot Demonstrator Project

**3.6** Once the research partnership as shown in Figure 1 was in place, the research process was adapted to ensure the integration of social science methodologies with technical development. To do this, the trajectory of the project was:

1. Establishing collaboration in and between the research team and research partnerships.

2. Developing research questions to underpin the interdisciplinary approach.
3. Training for the social scientists and technological development of the Grid.
4. Integration and analysis of the data sets.
5. Further Grid development and sharing outcomes among partners.

### ***Phase 1: Establishing Collaborative Research***

**3.7** The early fostering of the collaborative approach consisted of a range of meetings and discussions within the core team itself, and between the core team and the policy and practitioners group, and with the technical expert from the OGC. The main ethos of the project was that the research should drive the project and the needs of the research should shape technological developments. Each member of the team was aware that they needed to develop a coherent approach from the project's three main disciplines to undertake interdisciplinary research. There was also the extra complication of designing new research tools based on the Grid from the emerging interdisciplinary framework. The team decided to have a series of meetings (eight in total) in which each of the researchers discussed their own discipline's approach to the research problem, data sources, and theoretical perspectives.

**3.8** In these meetings the criminologists focused on whether the structure of a neighbourhood increases or reduces the likelihood that some people will commit offences or become victims of crime. The urban planners built on one of their 'needs analysis' projects that aggregated small area-level data across a range of education, health, and welfare domains to inform research and planning decisions regarding services for children and teenagers in Sheffield (Signoretta & Craglia, 2002). One of the sociologists developed research from his 'Communities that Care' project (France & Crow, 2005) to explore further the influence of neighbourhoods, schools, and family circumstances in relation to risk factors that may propel young people into crime. The Grid Officer participated in all the meetings to gain an understanding of the social science project in order to take an integrative approach in the development of Grid-based research tools.

**3.9** The project partners were an important dimension of the project in supplying data and support services. The core team had to develop a dialogue and shared understanding with this network of project-partners. Part of achieving this dialogue was the 'kick-off meeting' in which the core team met representatives from regional policy-makers, service and data providers, and technology providers (OGC and WRG) to discuss the project. The focus of the kick-off meeting was to outline the project, look at the kind of data needed, and introduce everyone to the Grid. The partners saw the project in terms of its social research potential and as an opportunity to support the emerging public policy research alliances in the region, rather than an explicit e-research project. The time and effort spent by the core team at this stage was important because it allowed each partner to see potential value in the project, rather than forcing each of them to subscribe to an assumed single shared objective. This helped to ensure that the pilot e-research project was shaped by the needs of research and researchers rather than being driven by the need to develop the technology.

### ***Phase 2: Developing Research Questions***

**3.10** After conducting literature reviews focusing on environmental criminology, Grid technology, and risk factors for young people, the team discussed how these perspectives could be integrated and how the Grid should be utilized to undertake large-scale data analysis over distributed information resources. The perspectives of the criminologist and urban planners with regard to the relationship between crime and neighbourhood contributed to generating several hypotheses but the team, and in particular the sociologists, had to find ways to encompass the area of young people at risk within a crime-neighbourhood dynamic. The sociologists were determined to develop a theoretical framework that would inform the interdisciplinary research and help to select data sets and research methodologies. They formulated the following research questions to aid the researchers in developing a multi-disciplinary framework:

1. Can we construct a reliable set of measures for community-based risk factors that allow us to measure them at ward or neighbourhood level?
2. Can we create a 'national norm' with which to compare?
3. If so, what relationship might exist between levels of risk and levels of crime — are neighbourhoods in communities with high-risk level young people also areas with high-risk factors?
4. What relationships might exist between levels of risk and levels of crime?

**3.11** These questions prompted the urban planners and the criminologists to consider their understanding of 'place' and 'crime' and to ask what data would be needed to address the questions. The criminologists argued for data that could link the locations of known offenders, offences, and victims. The urban planner wanted data that brought together lifestyles and socio-economic characteristics in clearly identifiable geographic units. The sociologists required data on young people between 16-18 years old — where they lived as well as their socio-economic circumstances. The researchers also felt they needed Census data at

Output Area (OA) and Super Output Area (SOA) as well as data from the Index of Multiple Deprivation to develop a robust model at the national level and to compare regional results with national findings.

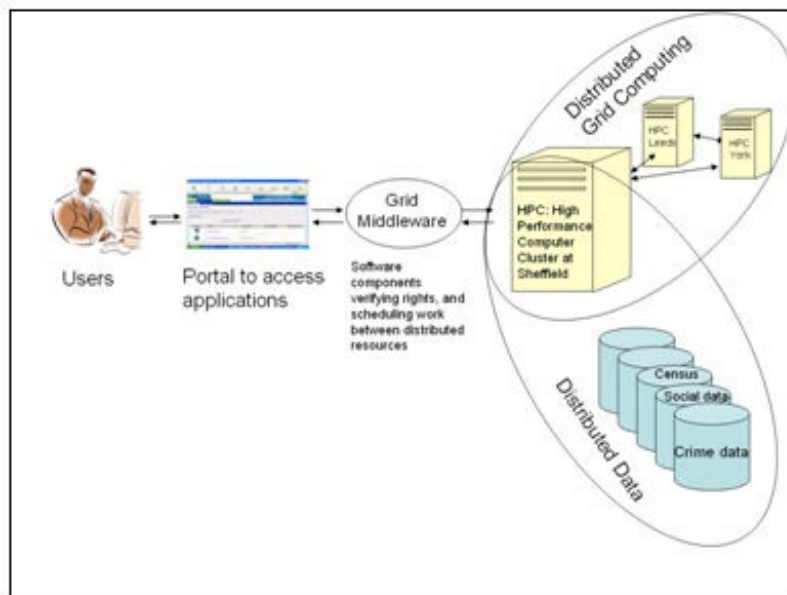
**3.12** Having observed the social scientists at work, the Grid Officer had to support the researchers in shaping the Grid for e-research. He knew that the WRG computing cluster provides a reliable service that could be assessed in a variety of ways with varying degrees of interactivity, but he felt that the researchers, who had little experience of high-powered computing, would find accessing the computing cluster technologically daunting. He explained that researchers in the WRG computing cluster are presented with terminals and expected to manage their user account through a command line using UNIX or LINUX, which are powerful operating systems. However, this means that researchers require additional training if they are only familiar with desktop environments. Once the social scientists had developed a framework of research questions and the Grid Officer had gained an understanding of what the social scientists were doing, the team undertook further research training in using the Grid for e-research and in specific statistical software packages.

### ***Phase 3: Training for the Social Scientists***

**3.13** Even though the social scientists were experienced researchers, the development of e-research meant that they needed to expand their skills. The Grid Officer showed the researchers how the Grid worked so that they could articulate what they needed from the tools. The researchers were given passwords and identifiers to work on the WRG. They learned to access the UNIX host 'Titania', work with files, manage directories, share data using UNIX groups, and access shared data and transfer files, and then progressed to working with scripts to try out the Sun Grid Engine. Furthermore, because the use of Grid computing resources was restricted to registered members of the academic community, it was also necessary to install secure shell client (SSH) security software to control the authorization levels of each user. The researchers therefore had to learn new concepts such as the 'tunnelling structures of communication', which means the way SSH encrypts communications so that they are secure over the Internet. The researchers had to learn about all these technological practices before they could start addressing the substantive research issues to develop e-research.

**3.14** Once the researchers had gained knowledge of some of the technological aspects of using the Grid, those without GIS experience had to learn about the statistical package, SPlus, and the geographical information system (GIS) ArcView. The researchers' training on the Grid enabled them to start using ArcView on the UNIX platform and to begin to work with the idea of a cluster system, rather than the more familiar PC environment. Another aspect of research-based learning involved the sociologists and criminologists learning about the visual logic of mapping in representing data. The PI, who was an urban planner, stepped in and talked them through what a 'view' was by explaining that: 'A view is composed of multiple maps, each with their own table'. Once the criminologists and sociologists could follow the logic of a thematic map and classification field, they could see how the maps worked as 'visual representations' of data. The Grid Officer also started to work closely with the PI so that the latter could understand how the mapping could work from 'a Grid point of view' and see what tools were needed for mapping with the Grid infrastructure.

**3.15** The Grid Officer continued to develop ways to facilitate Grid access. The configuration used at Sheffield involved a cluster of ten Sun machines connected to two other clusters at Leeds and York, forming the WRG (see Figure 2). The Grid Officer installed SPlus and ArcView software on the UNIX cluster. However, to ensure access to the distributed processing capabilities of the Grid, he had to write bespoke applications (specific applications for particular contexts of use). The first application he developed allowed researchers to retrieve Census data from the dedicated service MIMAS, at the University of Manchester, and store it locally to perform subsequent analysis. This first application tested the development toolkit of EASA (Enterprise Accessible Software Application), which had two benefits: (1) the researchers were happy because they had a highly usable interface, and (2) in EASA it was much easier to develop user interfaces for applications than using the traditional tools of the computer programming community.



**Figure 2.** System Architecture of the Project

**3.16** Figure 2 shows the architecture of a Grid such as the White Rose Grid, which was built by collaboration between the Universities of Sheffield, York and Leeds. The social scientists access the Grid via the EASA interface on their computers, then the technology (the 'middleware') mediates their access to the Grid (the 'Titania' cluster at Sheffield, and the remote ones at York and Leeds) that links them to data sources such as MIMAS and local data sets, as well as analysis tools such as S-plus and ArcView for processing the data.

#### ***Phase 4: Integration and Analysis of the Data Sets***

**3.17** The analysis of data was divided into three stages: (1) the team analyzed individual variables of the crime and youth data sets, including aggregation at census geography level, calculation of counts, rates, and standardized rates, and identified outliers and extreme values, (2) the researchers analyzed each variable in relation to key census variables and MOSAIC classification, and (3) they analyzed the key variables to identify statistically significant relationships, supported by their prior review of the literature on environmental criminology. Thus, for example, the criminologists and urban planners worked together on the unique crime data set provided by South Yorkshire Police (following a written protocol with the University of Sheffield), which included:

1. 371,000 reported victims of crime.
2. 46,800 offenders who have committed 118,000 offences.
3. 17,000 young offenders who have committed 45,000 offences.
4. 70,000 thefts from cars, 63,000 burglaries, 28,000 cases of damage to dwellings.

**3.18** The data were provided with X, Y coordinates at 10-metre resolution or better (1 metre in some instances) and covered the time period 1998-2003. This collection of data provided the opportunity to link the locations of known offenders, offences, and victims as required for the project. This had not been reported before in the literature because of the difficulty in accessing offender and victim data.

**3.19** The researchers then downloaded Census data at Output Area (OA) and Super Output Area (SOA) for the whole of England. They drew on 120 variables (based on the literature review) in the following domains:

1. Ethnicity and age
2. Economic activity and occupation
3. Socio-economic classification and qualifications
4. Household characteristics, including vacancies and overcrowding
5. Tenure
6. Car ownership
7. Migration

**3.20** This was used in conjunction with the Connexions South Yorkshire data for all 16-18 year-olds in South Yorkshire (approximately 30,000 young people) from November 2003 to March 2004, which included unit postcode, age, sex, ethnicity, and whether or not in education, employment, or training. Experian made the MOSAIC classification available, providing data on consumers for the whole of Great Britain at unit

postcode level (over 1.5 million records). MOSAIC classification segments consumers into 11 major groups and 61 detailed types, based on socio-economic characteristics and lifestyle surveys. The team also used the 2004 Index of Multiple Deprivation, containing seven domains which relate to: income deprivation, employment deprivation, health deprivation and disability, education, skills and training deprivation, barriers to housing and services, and living environment deprivation and crime at Super Output Area (SOA). The researchers undertook statistical modelling which involved two stages: (1) stepwise regressions to identify the most significant variables in accounting for the variance of offenders and to reflect on the findings in the light of the literature review, and (2) experimentation with different types of models to see which would yield the best results.

**3.21** On the basis of the analysis carried out for the county of South Yorkshire, the researchers found that:

1. Over 70% of the variance of victimization was accounted for by the proximity of the residential location of offenders.
2. Offences such as domestic burglary and criminal damage were also strongly related to the location of offenders.
3. The geographic distribution of offenders appeared to have strong correlations with Census data.

### ***Phase 5: Grid Development and Sharing Outcomes Among Partners***

**3.22** After developing a model of the distribution of offenders for South Yorkshire and validating it against the observed results, the model was extended to the whole of England at SOA and then filtered through a 1 hectare-cell grid based on the residential postcodes provided by the Royal Mail. The advantage of this procedure was that it reported the results of the model more accurately in relation to where people live. This was considered better than the system of large polygons that integrate sparsely-populated areas with more densely-populated ones. The final model selected was a General Linear Model of the Poisson family, in which the response variable (counts of offenders) was transformed using a logarithmic function. The model was then run for the whole of England at SOA in the following form:

$$\log(\text{observed/expected values}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6$$

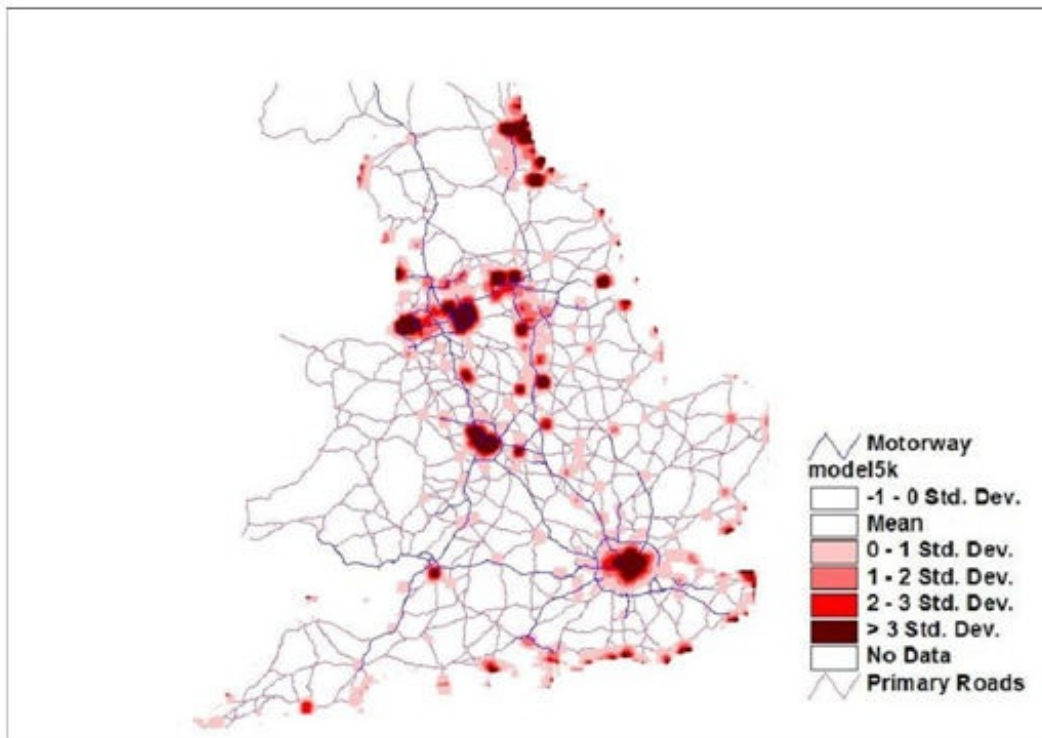
where:

X1 = Percent of economically-active unemployed  
X2 = Percent of households renting from another (hostels, secure accommodation, prisons, boarding houses, hotels, and other communal establishments)  
X3 = Percent of households with lone parents with dependent children  
X4 = Percent of residential spaces vacant  
X5 = Index of Multiple Deprivation 2004 Health Domain score  
X6 = Index of Multiple Deprivation 2004 Crime Domain score

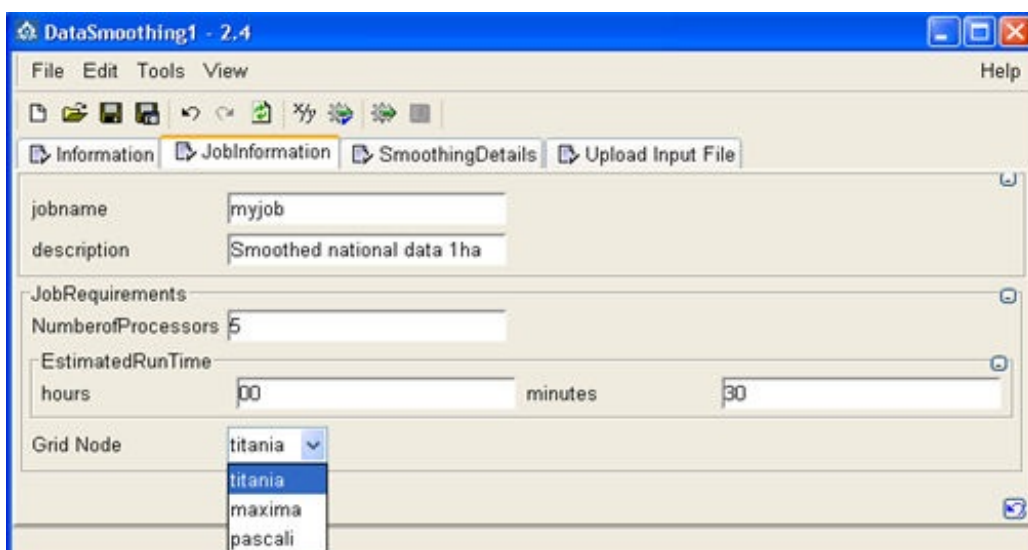
In order to explore the spatial patterns of the model, it became necessary to generalize (smooth) results at different scales.

**3.23** It was at this point in the project that the researchers needed the Grid to help them undertake smoothing at 5 km in an efficient manner. To this end, the Grid Officer developed the second main project application, which enabled smoothing of the model results for England at different scales. This was particularly important for the project, as smoothing the data at 5 km involved calculating the mean value across the neighbouring 50 cells for each of the 35 million hectare cells covering England and returning the result for display (see Figure 3). Figure 4 shows the portal developed by the Grid Officer for the researchers to access the application. The researchers, having imported an ASCII file with the data, could select the number of processors on which to operate. The portal would then schedule the operation via the Grid Middleware, based on the schema shown in Figure 2.





**Figure 3.** Modelled Distribution of Offenders in England at 5 km



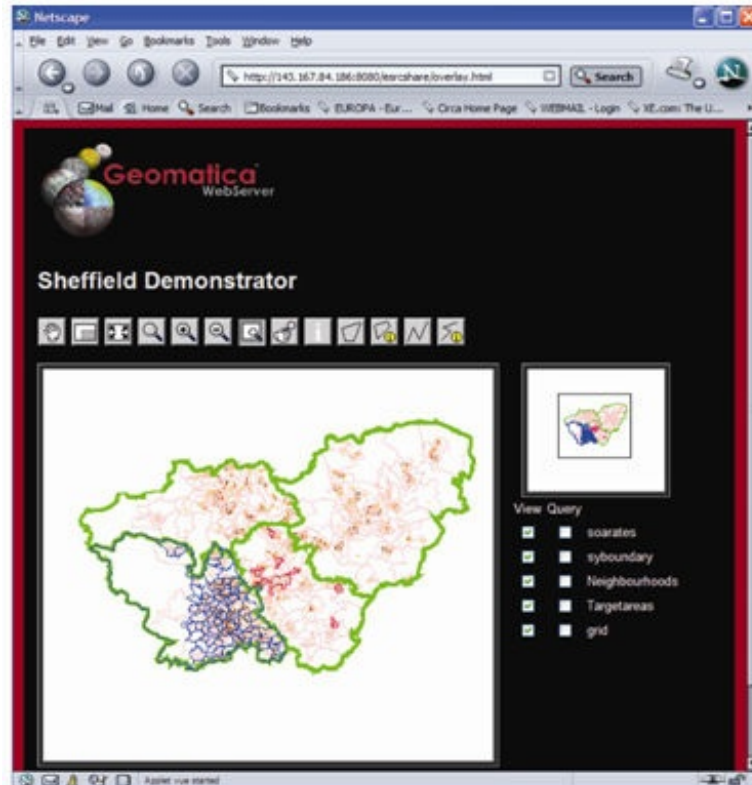
**Figure 4.** User Interaction with the Grid via the EASA Portal: Setting the Parameters for the Smoothing Process and Selecting the Number of Parallel Processors.

**3.24** It was at this stage that the team saw the advantages of using multiple high-performance computers – as smoothing at 5 kilometres took 6¼ hours on a single Pentium 4 processor (3 GHz) but only one hour when performed over 15 parallel processors on the White Rose Grid. This performance was achieved using the specific smoothing algorithm created in the project, because the internal routines of ArcView Spatial Analysis or ArcGIS required too long to compute the necessary calculations. Thus, at this stage of development, the team learned how useful Grid computing could be, provided that researchers are able to access the Grid infrastructure and parallelize the processing. However, doing so may require researchers to have the ability to access or write the appropriate algorithms, since most off-the-shelf software on a desktop cannot harness the Grid (Clematis, Mineter & Marciano, 2003).

**3.25** Having successfully processed the data through the WRG services, and re-imported the results into GIS for display purposes, the next issue was how to share the results obtained with other partners in the project, including South Yorkshire Police and the four local authorities in the region. In policy terms, the value of this data was that it enabled the comparison of areas having a relative higher level of risk as identified in the statistical model, with areas which are the subject of different policy interventions by agencies operating in the region who share responsibility for crime reduction strategies with the police. The



team particularly wanted to identify whether there were any gaps, i.e. areas at high risk not covered by targeted policies. The team installed Web Map Services (WMS) on the servers of the University and those of the project partners. These services, based on international standards, make it possible to overlay the maps of the model and its results held at the University with those of the policy boundaries and other data held by the partner organisations (see Figure 5). This solution allows each partner organisation to retain control of its own data, maintain it easily, and determine the level of detail and attributes to be shared. These are critical aspects of sharing information between organisations and it shows the value of spatial data infrastructure architecture to enable data sharing for policy purposes. These processes allow external partners to take advantage of the Grid's processing capabilities at the results level.



**Figure 5.** OGC-compliant Web Map Service Allowing Overlay and Query of Distributed Data

## **Discussion of Findings and their Significance in Developing e-Research**

**4.1** The knowledge that emerged from the project shows how significant the relationship between research practice and research tools is. By addressing the practices of social science research in the development of e-research we can start to understand and evaluate the development of the Grid in the e-research process. The description given in this paper shows that the development of research questions and their theoretical and methodological frameworks give shape to research projects and practices. Privileging social science knowledge and using that knowledge as a base for developing tools ensures that e-research can be shaped to meet the needs of social scientists. However, the project also shows that the development of new practices and tools requires additional support from data suppliers, project partners, and an emerging knowledge community. Furthermore, the development of e-research involves a learning process for researchers, technological suppliers and developers as well as stakeholder partners in projects. The need for new forms of support and ongoing learning means that time must be allocated in the research process for learning and development. However, if, as in the project described in this paper, the development of e-research is reflective and reflexive then e-research can enhance social science research.

**4.2** The quantitative findings of the project are important and confirm qualitative interviews with young offenders (Wiles & Costello, 2000). The findings based on the theoretical framework developed, the data sets and e-research tools used in the project all contribute to the theoretical and policy debates within environmental criminology. In particular, while offences are often correlated to indices of deprivation and their geographical distribution, evidence from this project indicates that the geographical link between offences and deprivation is not direct but 'mediated' by the geographical distribution of offenders. Furthermore, the strong link between Census data and offenders on the one hand, and offenders and victimization on the other, supported the case for modelling the geographical distribution of offenders on a national basis using the Census variables, which could be used to develop a relative 'risk' map for potential victimization. The key advantage of using the model developed in the project is that it could be done using nationally, and freely available, data (the Census and Index of Multiple Deprivation) without requiring the very sensitive and

unique data set that was made available to the researchers for the region of South Yorkshire. Thus, the process of e-research situated within the knowledge of social scientists contributed to existing academic knowledge and produced a powerful model and tool to map potential areas of risk.

**4.3** The work done in the project also contributed to the policy-making process by developing a tool to map and share information within the policy-making community. However, to enable this information sharing with partner agencies that did not have access to the University intranet, the research team had to find a solution that would allow the controlled publishing of the findings outside of the University firewall. This type of publishing also has to allow for the overlay of policy boundaries defined by the other partners within their respective institutional frameworks. Given that boundaries are policy-driven and therefore dynamic, it is necessary that these are regularly maintained by the 'owning' organizations and published from their own servers. This means that sharing data needs to be based on distributed servers communicating dynamically. The team therefore had to develop a solution for sharing information from distributed sources. The solution adopted is based on the Web Map Service (WMS) specifications of the Open Geospatial Consortium (2001). A WMS is an interface for requesting map-based images from one or more distributed databases of geographic data. The response to a request is one or more geo-registered map images that can be displayed and overlaid in a browser application and is thus kept up to date and can be shared amongst partners. This type of Web service is now an international standard and a key component in the development of Spatial Data Infrastructures.

**4.4** When considering the substantive issues in developing e-research for the social sciences, it is evident that the development of new tools can enhance research but also require various types of work. Key aspects of creating e-research for the social sciences include the need for a collaborative network of research partners and a supportive learning environment for researchers. Furthermore, if e-research is to meet the diverse and rich research needs of social sciences, then its development must be based on the researchers' practices. The pilot demonstrator shows how the interaction between research practice and the development of new ICT research tools was influential in shaping the research process that led to the outcomes of the project.

## **Conclusion**

**5.1** The e-research demonstrator project 'Collaborative Analysis of Offenders' Personal and Area-based Social Exclusion' shows how a group of social scientists developed e-research, with the support of a Grid technologist. This paper shows that researchers' practices and their understanding of the theory and methodology of research were important in shaping e-research. In addition, the researchers needed a network of public data providers, commercial data providers and regional service data providers in order to undertake their research and analysis. The social scientists found that electronic access to distributed data sets and the Grid's computing power both enhanced their research, and that the development of visualisation to represent the findings benefited their research. The demonstrator was also useful to the policy-making community. The project therefore developed e-research in the social sciences – it used and shaped all three dimensions of e-research: (1) electronic distributed networks of data and research partners, (2) the use of high-powered computing for data analysis, and (3) visualisation. The team felt that e-research could be adapted for use in the social sciences as long as new research tools are shaped by the needs of the research.

**5.2** One of the lessons learnt at this early stage of the innovation and development of e-research for the social sciences is that time must be allocated to generating understanding between social researchers from different disciplines in developing interdisciplinary research. Another important lesson in developing e-research is that it requires a network to be built to support it and that network has to be built on trusted relationships. Finally, the research and research questions should shape the development of e-research and its tools – if the development is research-driven then e-research can be adapted to the needs of social science: technically-driven research will not adapt as easily to the needs of social science.

**5.3** To conclude, the pilot demonstrator project has contributed to theoretical debates on environmental criminology, to the development of user-focused training opportunities and to facilities that make Grid computing in e-research easier for non-experts. The development work produced a technical and social infrastructure to generalize operational data on crime and youths that can be served to researchers for modelling and (accessible) mapping, allowing local information to be used to address social issues which can inform policy at all levels of government. However, the project also revealed the existence of technical barriers that could only be overcome through *ad hoc* developments made by the project participants and this is an area that must be considered in future e-research in the social sciences. Moreover, further research is needed to understand the dynamics of producing social science knowledge through research practice, in particular the nature of interdisciplinary research and analysis within the development of e-research science.

---

## Notes

<sup>1</sup>The project was carried out at the University of Sheffield in 2003-04 with ESRC award RES-149-25-0027.

<sup>2</sup>Director General of Research Councils, Office of Science and Technology (UK) cited in *e-Social Science News Issue 1*, Summer 2005 published by National Centre for e-Social Science (NceSS), Manchester, UK.

---

## References

BEAULIEU, A. and WOUTERS, P. (2009) 'e-Research as Intervention' in JANKOWSKI, N. (editor) (2009) *E-research: transformation in scholarly practice*, New York: Routledge, pp. 54-69.

CLEMATIS, A., MINETER, M. and MARCIANO, R. (2003) 'High performance computing with geographical data', *Parallel Computing*, vol. 29, no. 10, pp. 1275-1279. [doi:10.1016/j.parco.2003.07.001]

CRAGLIA, M., WESSELS, B., GRIFFITHS, M. and COSTELLO, A. (2005) 'Building bridges between social science, grid, and geospatial communities: a reflection on practice', *Proceedings of the 1st e-Social Science Conference*, Manchester 22-24 June 2005.

FRANCE, A. and CROW, A. (2005) 'Using the risk factor paradigm in prevention: Lessons from the evaluation of Communities that Care', *Children and Society*, 19 pp. 172-183. [doi:10.1002/chi.866]

FRIEDRICH, J. and BLASIUS, J. (2003) 'Social Norms in Distressed Neighbourhoods: Testing the Wilson Hypothesis', *Housing Studies* 18(6), pp. 807-826. [doi:10.1080/0267303032000135447]

HALFPENNY, P., PROCTOR, R., YU-WEI, L., and VOSS, A. (2009) 'Developing the UK-based e-Social Science Research Program' in JANKOWSKI, N. (editor) (2009) *E-research: transformation in scholarly practice*, New York: Routledge, pp. 73-90.

JANKOWSKI, N. (editor) (2009) *E-research: transformation in scholarly practice*, New York: Routledge.

NENTWICH, M. (2003) *Cyberscience: Research in the age of the Internet*. Vienna: Austrian Academy of Sciences Press.

SAMPSON, R.J., RAUDENBUSH, S.W. and EARLS, F. (1997) 'Neighbourhoods and Violent Crime: A Multi-Level Study of Collective Efficacy', *Science* 277, pp. 918-924. [doi:10.1126/science.277.5328.918]

SIGNORETTA, P. and CRAGLIA, M. (2002) 'Joined-up government in practice: a case study of children's needs in Sheffield', *Local Government Studies*, vol. 28, no. 1 pp. 59-76.

WESSELS, B. and CRAGLIA, M. (2007) 'Situated Innovation of e-Social Science: integrating Infrastructure, Collaboration, and Knowledge in Developing e-Social Science', *Journal of Computer Mediated Communication*, 12 (2), article 18, <<http://jcmc.indiana.edu/vol12/issue2/wessels.html>>. [doi:10.1111/j.1083-6101.2007.00345.x]

WILES, P. and COSTELLO, A. (2000) *The Road to Nowhere: The Evidence for Travelling Criminals*, Home Office Research Study 207. London: Home Office.

WILSON, W.J. (1987) *The Truly Disadvantaged*. Chicago: Chicago University Press.

WOOLGAR, S. (2003) *Social Shaping Perspectives on e-Science and e-Social Science: the case for research support*, A consultative study for the Economic and Social Research Council (ESRC).