

THE NATIONAL STRATEGY FOR DATA RESOURCES FOR

Research in the Social Sciences



“Many of the big challenges confronting the UK – such as how to succeed in the global economy or the nation’s rapidly changing demography – are intrinsically social science in nature. Others, such as climate change or preventative health, addressed effectively, need social science solutions.”

ECONOMIC AND SOCIAL RESEARCH COUNCIL (2004)
ESRC CONSULTATION WITH STAKEHOLDERS: A SUMMARY
SWINDON: ESRC



PREFACE

Preface

We are now living in an increasingly complex, dynamic and diverse society. This means that there is a pressing need to create better resources to answer some of the more complex research and policy questions this poses. Developments in technology, particularly e-social science, are creating path breaking new opportunities to link, model and mine large scale datasets.

If the UK is to maintain its leading position there needs to be a commitment to the long term support of existing datasets and a radical extension of the current data infrastructure. We need to create bigger and better datasets with large enough samples sizes to enable social scientists adequately to answer a range of increasingly complex questions.

It is for that reason that the ESRC led the establishment of the UK Data Forum, chaired by Professor David Rhind. The UK Data Forum brings together representatives of the major government departments, research councils, the devolved administrations and other organisations involved in funding and development of data collection activities.

I would like to thank Professor Rhind and members of the Forum for their hard work in preparing this first version of the *National Strategy for Data Resources for Research in the Social Sciences*. I look forward to working with the Forum to implement the strategy that will ensure that UK social researchers are in a position to meet the challenges posed by 21st century society.



A handwritten signature in black ink that reads "Ian Diamond".

PROFESSOR IAN DIAMOND FBA, AcSS
CHIEF EXECUTIVE, ECONOMIC AND SOCIAL RESEARCH COUNCIL

01	FOREWORD
02	EXECUTIVE SUMMARY
05	INTRODUCTION AND AIMS
05	1.0 THE NEED FOR A STRATEGIC APPROACH TO THE DEVELOPMENT OF DATA INFRASTRUCTURE FOR SOCIAL RESEARCH
06	2.0 THE SCOPE OF THE NATIONAL DATA STRATEGY
07	3.0 SHAPING THE STRATEGY – LINKING FUTURE RESEARCH CHALLENGES TO DATA RESOURCES
11	4.0 KEY ACTIONS TO INITIATE AND PROMOTE THE NATIONAL DATA STRATEGY, 2006-07
17	5.0 MAKING THE NATIONAL DATA STRATEGY WORK
21	APPENDIX
23	ANNEX

FOREWORD

Foreword

01 <

Progress in research in the social sciences, as in the natural sciences, is dependent upon appropriate data. Whether these arise through monitoring the activities of people, recording values and beliefs of individuals or groups, or measuring the performance of organisations, they provide the basic information through which researchers seek to understand the structure of and changes in society.

Over the past decade or so, our capacity to create, record, store and retrieve data has grown rapidly as information and communication technologies become pervasive. In line with this rapid growth in the volume of data now potentially available for research, the research community has increased its demand for more sophisticated approaches to the creation and use of such data. Censuses, longitudinal surveys and administrative data all now form part of the range of resources that social scientists need to access and explore for research purposes. Moreover, linking different data sets together can give fresh insights.

As a consequence of these trends, we are now in a position where the scope to provide data for research purposes has extended, but – inevitably – the resources to develop and manage data for research remain limited. And it often takes years to assemble data relevant to new issues. For these reasons, a more strategic approach to the development of data resources has been identified by the major UK organisations responsible for funding the creation and development of social science data. This approach, described in this publication, starts from an assessment of likely future research issues, identifies the actions required to ensure that relevant data to address these issues will be available in a timely manner, and then prioritises these actions in line with available resources. Perhaps uniquely, the creation of the National Data Strategy has involved a very active collaboration between research councils, Whitehall government departments and the devolved administrations, charitable trusts, academics and the Statistics Users Forum. Whilst there are differences in the time horizon over which the different bodies operate and some differences of emphasis, there was a remarkable degree of consistency between them on likely future research priorities.

Creating the plan set out in this document and implementing it is valuable, but it only represents the start of what must become a more continuous process. While it remains complicated by the range of resources that need to be coordinated, the number and different perspectives of the different agencies involved in these developments and the tension between long range research plans and the shorter term need to inform policy developments, all are agreed that further review and updating is essential.

I am hugely grateful to all those who took part in the creation of this first National Data Strategy. All members of the UK Data Forum (who are identified in the Annex, page 23) brought expertise, insight and enthusiasm to the task. Most of all however, I am grateful to Peter Elias who steered the work and did much of it himself. Finally, I am delighted to acknowledge the key role played by the ESRC in funding and otherwise facilitating the whole project.



David Rhind

DAVID RHIND CBE, FRS, FBA
CHAIR, UK DATA FORUM

› 02 Executive Summary

Research across the social sciences contributes greatly to our knowledge about the way we live. It informs our understanding of the links between social and economic well-being, the longer term impacts of choices and decisions we and others make as children, students, partners, parents and citizens, and of the ways in which we relate to each other. Much of the evidence upon which social and economic policies are based derives from research conducted by social researchers: economists, sociologists, psychologists, geographers, political scientists and others working across the range of disciplines which constitute the social sciences.

The types of information that social researchers draw upon to provide this evidence are varied. Some comes from small scale observational studies. Increasingly though, large scale national resources are used for this purpose. These include national census and survey data sources, or data which are generated via administrative records, such as hospital visits, social benefits received or schools attended. Census, survey and administrative data, when suitably organised, stored, documented and made accessible for research purposes, form part of what is termed the *national infrastructure for social research*.

Developing this national data infrastructure, to ensure that relevant information is available for research at the appropriate time, requires a degree of planning and foresight. Some data resources, particularly those which are based upon repeated surveys of the same individuals or organisations (longitudinal data resources) may take many years to reach the stage where they can provide valuable information for research purposes. Other resources may already exist, but access to them is restricted for legal or ethical reasons. Some can be created via data linkage, but the resources required to achieve this are difficult to mobilise. Some may not exist at present and must be planned, resourced and developed.

The *National Strategy for Data Resources for Research in the Social Sciences* (henceforth the National Data Strategy) is a plan which aims to ensure that the national data infrastructure meets the demands which will be placed upon it to address both current and future research needs. It has been created by the UK Data Forum, a body which brings together those with responsibilities for the development of data resources across two research councils, research foundations, a range of government departments, the Office for National Statistics (ONS) and the devolved administrations. It builds upon work undertaken by many of these bodies to identify the key research challenges they will face over the next 5-10 years and sets out a series of goals that must be met in the short term to provide the resources required to address the longer term research issues it has identified.

The National Data Strategy is not set in stone. Each year the UK Data Forum will revisit this plan, both to monitor its progress and to re-evaluate research priorities as appropriate. Four key research challenges have currently been identified as areas which are likely to require a wide range of data from a variety of sources. These are: *the ageing*

population, migration, globalisation and child development. While there are numerous other topics that will benefit from social research over this period, there is general agreement on the need for significant research effort to be focused on these challenges. Each is likely to give rise to concerted scientific investigation over the next 5 to 10 years if we are to improve our understanding of recent and forthcoming trends and developments, and to inform policies designed to address issues arising in these areas.

The data resources required for this purpose share the following characteristics:

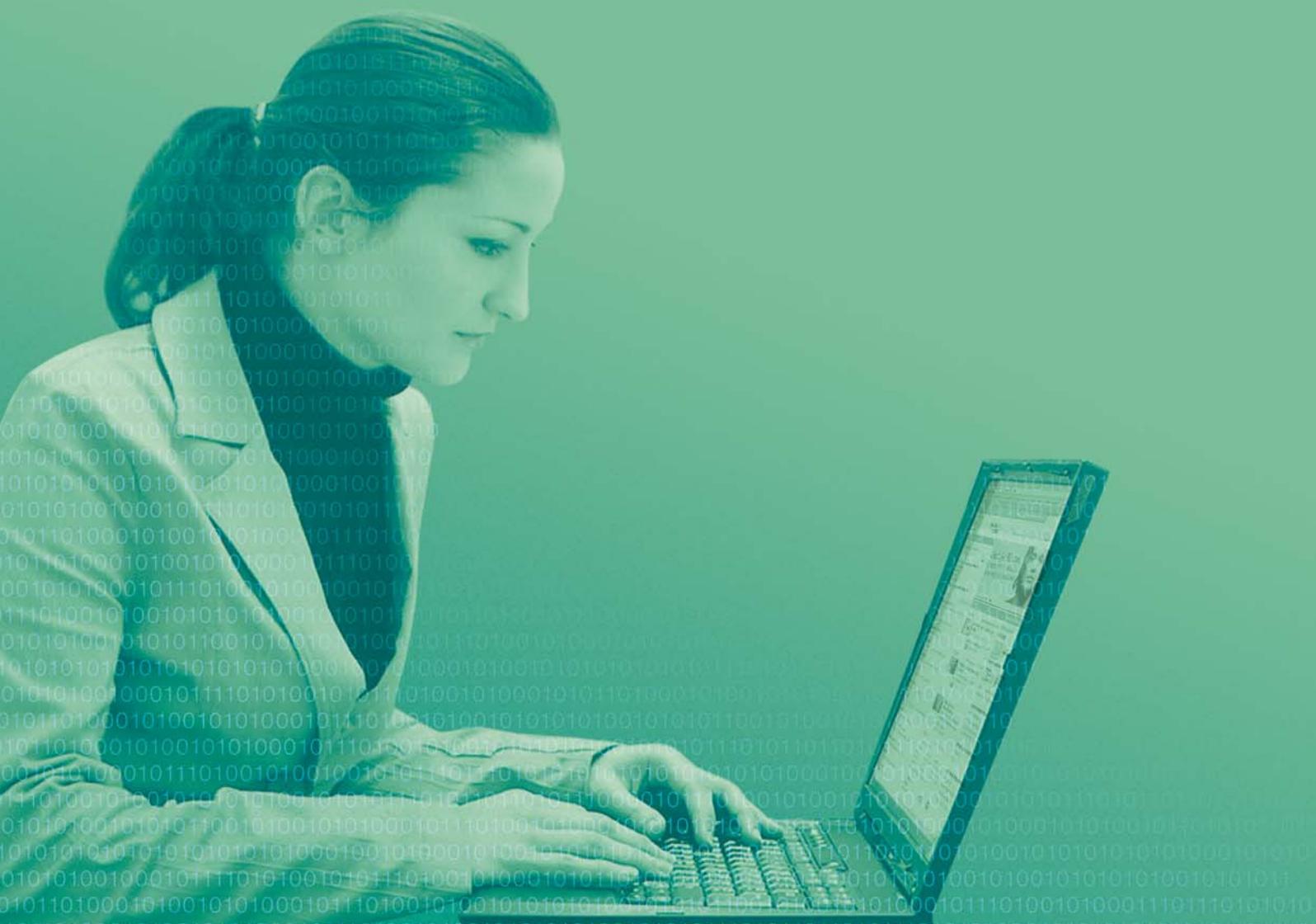
- They should measure/monitor individuals/organisations at the most detailed level and should be available for research purposes at this 'micro' level.
- They should facilitate the monitoring of trends or transitions (eg via repeated surveys or longitudinal surveys).
- They should facilitate linkage between different sources of information, enabling researchers to enrich and extend their studies into areas such as health, genetics, and environmental change.
- They should be available across diverse populations, facilitating the study of interesting subgroups within those populations and allowing researchers to make use of the natural variation within and between populations.

The National Data Strategy identifies the types of data resources that fulfil these requirements, either in whole or in part, and specifies a set of actions that will safeguard these resources and/or improve their use for research purposes. These include:

- The Economic and Social Research Council (ESRC), the ONS and the devolved administrations working closely to shape plans for access to future census and survey data.
- The ESRC, the Medical Research Council (MRC) and the Wellcome Trust working together to improve access to longitudinal survey data that can inform research at the boundary between the medical and social sciences.
- A range of government departments contributing to an exploration of the potential that many administrative data sources have to enrich existing data sources via data sharing and data linkage.
- Following a review of longitudinal data resources, the ESRC working with other potential funding agencies (government departments, research foundations) to ensure that the resources required to strengthen such resources are forthcoming.
- The Department for International Development working with the ESRC, Eurostat, international organisations responsible for the collection of data resources and other national funding agencies – co-ordinating the development of a wide range of resources to facilitate detailed social and economic research on an international scale.

“The Government is currently looking into how to co-ordinate better research and analysis, particularly where the issues are cross-cutting. Issues relating to data sets will be considered as part of this work.”

THE WORK OF THE ECONOMIC AND SOCIAL RESEARCH COUNCIL:
GOVERNMENT RESPONSE TO THE COMMITTEE'S FIRST REPORT
OF SESSION 2004-05
FIFTH SPECIAL REPORT OF SESSION 2004-05



1.0 INTRODUCTION AND AIMS

- 1.1 National data resources for social scientific research – measuring and recording information about people, families, places, organisations, households, and the relationships between them – form an important part of our research infrastructure. This infrastructure is generated in a number of ways, via data collected in censuses, from surveys and through administrative record-keeping. The data infrastructure so created is a vital research resource, informing research issues not just within the social sciences but also at the boundaries between social science and other scientific disciplines, and helping governments make effective policy decisions based on the best available evidence.
- 1.2 This National Data Strategy aims to provide a coherent framework for the development and maintenance of a robust data infrastructure, ensuring that relevant and timely data are available to inform and address future research priorities.
- 1.3 In meeting this aim, key elements of the National Data Strategy are that it should be flexible and responsive. Issues presently regarded as future research priorities will change and evolve as time progresses. The National Data Strategy will be reassessed annually, linking evolving research issues to future data needs, identifying gaps in data provision and co-ordinating plans to address such gaps.

2.0 THE NEED FOR A STRATEGIC APPROACH TO THE DEVELOPMENT OF DATA INFRASTRUCTURE FOR SOCIAL RESEARCH

- 2.1 The United Kingdom has a stock of world class data resources which informs and underpins much research across the social sciences and which complements those held in other countries. This, in turn, helps recruit and retain leading researchers in the social sciences and related fields. A key task is to look ahead to the next generation of UK resources in order to sustain this leadership while improving connectivity with similar resources elsewhere.
- 2.2 The existing portfolio of resources includes *national census and survey data*, most of which are financed and developed by the major spending departments of government and collected in collaboration with the ONS. Among such resources, *longitudinal survey data*, derived from repeated measurements taken from the same units of observation (eg individuals, organisations, households) are particularly important, in that they allow monitoring of processes of change and facilitate research which investigates causal interpretations of these processes. Some major longitudinal survey data sources are funded directly by the research councils and charitable trusts (eg ESRC and/or MRC, Wellcome Trust) often with government departments and agencies contributing to the costs of data collection. Some are funded directly by relevant government departments and public agencies.
- 2.3 Other data resources with potential to inform social scientific research derive from *administrative and data logging systems*, recording transactions, registrations or monitoring information (traffic flows, pollution levels, etc.). Administrative data are held by both public and private sector organisations but remain, as yet, relatively underutilised as major national research resources.
- 2.4 There are significant costs associated with the maintenance and/or development of national data resources, particularly the population census and longitudinal surveys. Plans to enhance or add to such resources must be prioritised in terms of the need for research evidence. Cost sharing arrangements between funding

agencies might be appropriate, yet these are often difficult to plan and manage. Access to data for research purposes may sometimes be restricted for legal and/or ethical reasons. Data linking procedures are emerging as ways of enhancing existing resources or creating new resources, but these raise issues relating to consent, confidentiality and data security, and pose questions about the technical feasibility of linkage, the quality of the data resources so created and their suitability as research resources.

2.5 To tackle these issues, a more strategic approach is required through which all the bodies engaged in the production and maintenance of data resources can work together to resolve problems and to ensure that relevant resources are in place and ready for the key research challenges which lie ahead. The National Data Strategy is designed to provide this approach for social research.

2.6 The National Data Strategy sets out a plan for co-operation between the bodies engaged in the production and maintenance of data resources for social research. There are three parts to this plan: its scope (section 3); the issues which shape the strategy over the long term and the types of data resources that must be promoted to inform research on these issues (section 4); and the actions that will be undertaken over the next year to move towards these long run goals (section 5). The final section describes the mechanisms that have been put in place to deliver the National Data Strategy.

3.0 THE SCOPE OF THE NATIONAL DATA STRATEGY

3.1 In establishing the scope of the National Data Strategy, an important issue relates to the definition of 'social research'. Traditionally, this has been regarded as research which can be positioned within or linked to a particular discipline, for example: economics, sociology, political science, human geography. Increasingly, social research is defined not by the disciplinary affiliations of those who conduct research, but by the areas of knowledge and expertise which contribute, both theoretically and empirically, to the issues which are being addressed. This, in turn, has encouraged more cross-disciplinary working, not just across the traditional social science disciplines, but between the social sciences, natural sciences and medical science.

3.2 Many government departments and agencies, including the devolved bodies, also have a strong need for information which crosses these disciplinary boundaries. Their needs derive from the desire to plan, deliver and assess the effectiveness of the services they deliver.

3.3 To facilitate such research, partnership arrangements now exist between research councils and with government departments to facilitate and promote such co-operation – stimulating research which is breaking down the traditional barriers between disciplines and bringing into focus the need for a social dimension on many key scientific research issues. These arrangements now need to encompass the development of the data resources required for cross-disciplinary research.

3.4 The National Data Strategy for Social Research is a plan for the development of data resources by all relevant UK agencies. It involves partnerships between and contributions from the major research councils (principally ESRC, MRC and the National Environment Research Council (NERC), all UK government departments, regional and local agencies, the devolved administrations and the major research foundations. Whilst focused primarily within the countries of the UK, the strategy is not limited by national interests. An important feature of the National Data Strategy is that it should help promote and widen the scope for social research at the international level, through the development of access to data and via research collaborations beyond the UK.

4.0 SHAPING THE STRATEGY – LINKING FUTURE RESEARCH CHALLENGES TO DATA RESOURCES

- 4.1 Plans to collect and develop data need to be informed by a view of future research activities. Given that many of the large scale surveys and databases which constitute the knowledge base for research take a considerable time to reach ‘maturity’ – the point when they start to yield valuable information for research purposes – the strategy is shaped by views of the research challenges which lie 5 to 10 years ahead. This section describes how future research issues are identified and lists those which are currently emerging as of particular importance across the social science research community. A typology of data resources is used to indicate the link required between future research challenges and efforts to prioritise developments.
- 4.2 A variety of mechanisms exist within and between government departments to establish future needs for research evidence. The Research Councils, professional bodies and the learned societies also regularly undertake consultations with their communities of interest to develop an informed consensus about the topics that are most likely to require future research efforts. The appendix provides further details about these mechanisms.
- 4.3 Through these processes of consultation with academics, the learned societies and between government departments and agencies (national and international), a number of research priorities have been identified as areas within which the social sciences will make important contributions over the next 5 to 10 years. These include: the ageing population; migration; issues relating to child development and the challenges posed by the rapid pace of social and economic development in various parts of the world.
- 4.4 This list is neither exclusive nor exhaustive. Many important issues constitute major challenges not just for policy development but in terms of our understanding of their social and economic causes and consequences. These include environmental change, criminal behaviour, drug abuse, security and transport. The research priorities outlined below are highlighted for two reasons. First, they are broad in terms of the issues they pose and the new knowledge that needs to be generated to address these issues. A wide array of social science and other disciplines can contribute both theoretically and in terms of shared knowledge to an understanding of issues within each area. Second, they are cross-cutting in terms of the funding activities of the research councils and the policy interests of government departments, requiring collaboration between these bodies in identifying future information needs.

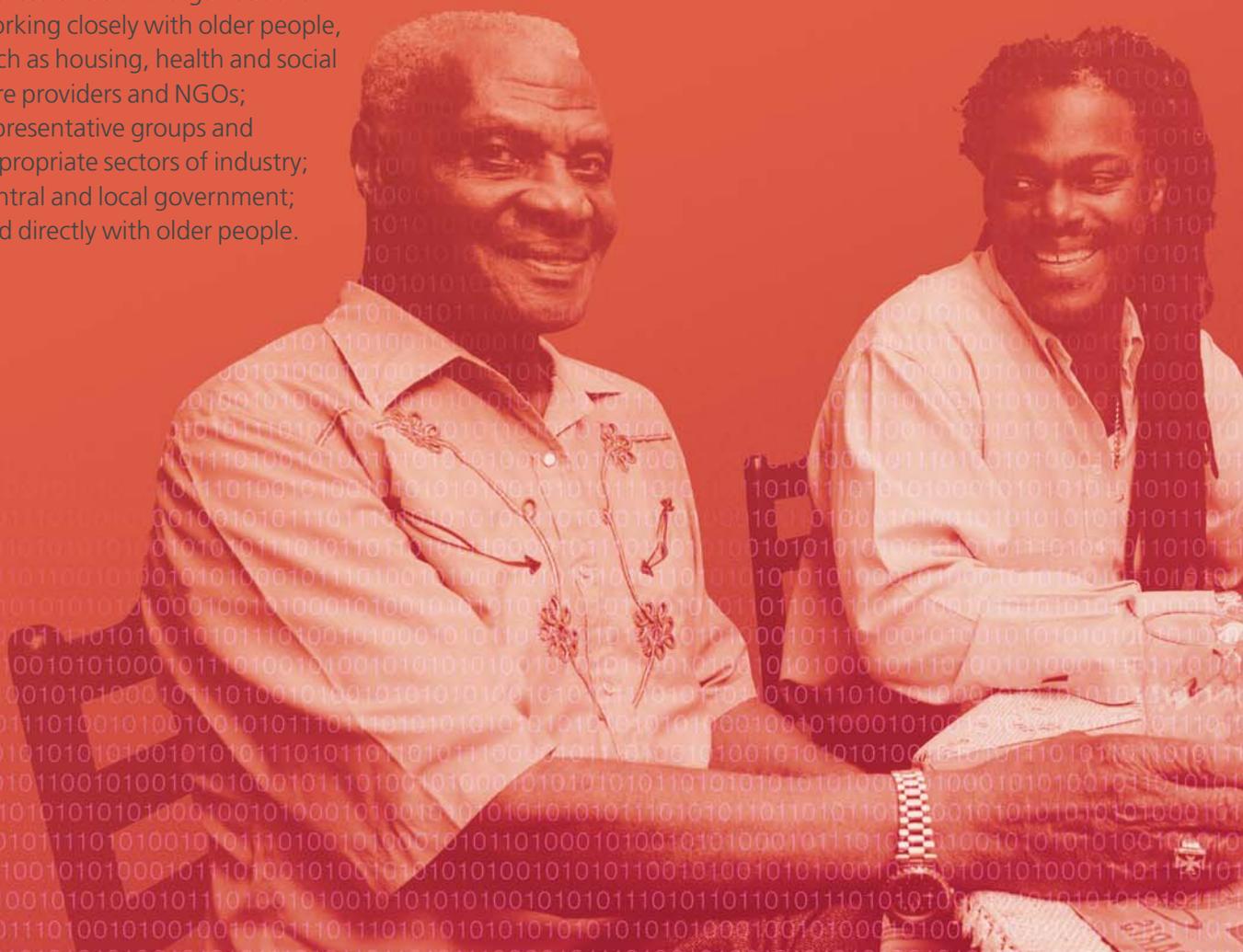
The ageing population

- 4.5 There is now strong evidence that, instead of reaching a plateau, life expectancy will increase significantly over the next fifteen years and beyond as mortality rates continue to decline among the older age groups¹. The problems posed by these demographic changes, – the ‘ageing population issues’ as they are more generally known – are of two types. First, there is a need to improve our understanding of the long term relationships between childhood development, lifestyle (eg diet, exercise, and community life), occupational health, morbidity and mortality. How is lifestyle related to health and how are related changes in behaviour brought about? Second, the health of an ageing population will determine the demands that will be placed on health systems, on social care, on pensions and public/private savings, and on the varying nature of consumers’ expenditures. Without the ability to anticipate many of these demands, infrastructure and economic plans to meet the needs of the population are placed at risk.

¹ House of Lords Science and Technology Committee (2005) *Ageing: Scientific Aspects*. Volume 1 Report. London: The Stationery Office.

The new dynamics of ageing – an interdisciplinary research programme

Building on their existing successful collaboration and previous initiatives on ageing, four research councils (ESRC, EPSRC, BBSRC and MRC) invited proposals under this major new interdisciplinary research programme on ageing. The programme is designed to bring together scientists from the full range of disciplines covered by the four Research Councils to investigate many of the longer term issues associated with an ageing population. The programme will also have a strong user focus through collaboration with those professionals and organisations working closely with older people, such as housing, health and social care providers and NGOs; representative groups and appropriate sectors of industry; central and local government; and directly with older people.



- 4.6 These processes are common in many countries, though there are significant differences between and within countries. An important aspect of the data resources required to study issues related to ageing is the ability to conduct intra- and international comparative research.

Migration

- 4.7 Related to the issue of demographic change, varying patterns of migration both within and between the countries of the UK, from other European Union countries and from/to the rest of the world will impact upon infrastructure planning in a variety of areas. The delivery of better housing, health facilities, schools and transport networks may be both a cause and a consequence of migration. Added to this complexity is the possibility of significant inflows of migrant workers in future years to support an ageing population – and the consequences of such migration in terms of future population dynamics. Understanding and anticipating the social, cultural and economic impacts of varying migrant flows, including understanding the role of developments in countries migrants come from, present major challenges for social researchers.

Child development

- 4.8 Changing patterns of family formation and dissolution are having a profound effect upon the nature of the developmental environment for children. Concerns about the relationship between these trends and ‘early years’ socialisation, the physical and psychological well-being of children, and the role of personal and social education have been and will remain as important issues for social research. The aim to eliminate child poverty and reduce inequalities in incomes and access to resources lends impetus to the need to improve our understanding of the long term relationships between genetic endowments, childhood experiences, education, adult behaviours and health.

Changing global economy

- 4.9 While the UK has long experienced the impacts of shifts in the balance of power and global trading patterns, the acceleration of recent trends which look set to position countries such as China, India and Brazil at the centre of global economic growth will have major implications for the UK. How will the UK adjust to new trading patterns? What are the implications in terms of political power and influence? Related to these changes is the issue of sustainable development and climate shift. There is now general agreement among the scientific community about the causes of the latter, yet the consequences are as yet, unpredictable. Equally, knowledge about the social impacts of macro and micro approaches to sustainable development is limited. Social and economic research will play an important role in helping to turn scientific knowledge into practical and attainable policies to address such issues.

- 4.10 The data resources required to address these key challenges are varied. Six broad types of resources are considered in relationship to the research challenges. These are:

- **Census and population surveys** – examples include the decennial Census of Population, the Annual Census of Employment and the proposed new Integrated Household/Continuous Population Survey, all of which provide vital information about population change, population diversity, economic activity and yield the detailed spatial information required for small area statistical needs.

10

- **Administrative data** – a wide range of statistical information arises from electronic transactions (eg payments, purchases, registrations, licensing, taxation, treatments) and remote monitoring (eg traffic flows, pollution monitoring, climate recording, web hits). Data so generated can often be linked in some manner – by names and birthdates, business and residential addresses or other identifiers (eg National Insurance numbers, National Health numbers, student numbers, etc.). Linking may be achieved through time, thereby creating longitudinal administrative records, by location to provide more detailed geographical information, or both.
- **Longitudinal surveys** – the British Household Panel Study, the British Birth Cohort Studies, the Census Longitudinal Studies, the Workplace Employee Relations Surveys and a wide range of other longitudinal surveys provide detailed information about processes of change, often over long periods of time for organisations, households or individuals. The importance of such sources relates to their ability to capture information on transitions and to relate these to longer term outcomes.
- **Socio-medical data** – resources which combine both detailed biomedical information with socio-economic data from individuals, some of which are also longitudinal in nature. Recent additions to the range of data available for research at the intersection of social, economic and biomedical study include sources such as the biomedical sweep of the National Child Development Study and the English Longitudinal Study of Ageing.
- **Business and economic data** – information collected by or from organisations, often for commercial purpose. These include loyalty card data from major retailers, customer service records, detailed financial and other business information required for national accounting purposes, and surveys of workplaces for a wide variety of purposes (eg Annual Survey of Hours and Earnings, Workplace Employee Relations Survey).
- **International data** – many other countries collect data equivalent to the types listed above. Additionally, some data resources have been developed specifically to facilitate comparative research or for supra-national requirements (eg European Social Survey, European Labour Force Survey, European Union Survey of Income and Living Conditions). For comparative research purposes and to facilitate international collaborative research efforts, such resources are of major importance.

Table 1: Linking research issues to data structures

Data structures/ Issues	Census and population surveys	Administrative data	Longitudinal surveys	Socio-medical data	Business and economic data	International macro/micro data
Ageing population	✓✓	✓✓	✓✓	✓✓	✓	✓✓
Migration	✓✓	✓✓	✓	✓	✓	✓✓
Globalisation	✓	✓	✓	✓	✓✓	✓✓
Childhood development	✓	✓	✓✓	✓✓	✓	✓✓

Key: ✓ = Relevant data resource ✓✓ = Critical data resource

- 4.11 For each type, Table 1 links the research issues outlined earlier in this section with the range of research resources available for research purposes and indicates the potential such data have to inform future research on these issues.
- 4.12 A number of resources stand out as major sources of information to inform future research across a range of issues. These are: census and population surveys (eg the Census of Population 2011 and the proposed Integrated Population Survey), administrative data, longitudinal surveys (panel, birth and age cohort studies) and international counterparts to these resources.
- 4.13 It is equally important to recognise that, without the skills and knowledge to analyse and interpret data, and in the absence of good access arrangements and relevant information about the ways in which data are constructed or derived, the full potential that various data resources have to inform research issues will not be realised. The development of data resources must be accompanied by relevant training in their use, with efforts to improve data access and by working to improve and standardise documentation.

5.0 KEY ACTIONS TO INITIATE AND PROMOTE THE NATIONAL DATA STRATEGY, 2006-07

- 5.1 This section outlines activities that are currently underway or are planned for the period April 2006 to March 2007 to assist with the development of data resources for the research challenges that lie ahead. Each year the National Data Strategy will be updated to take stock of progress towards its strategic priorities, to reassess these priorities and to initiate new activities as appropriate.

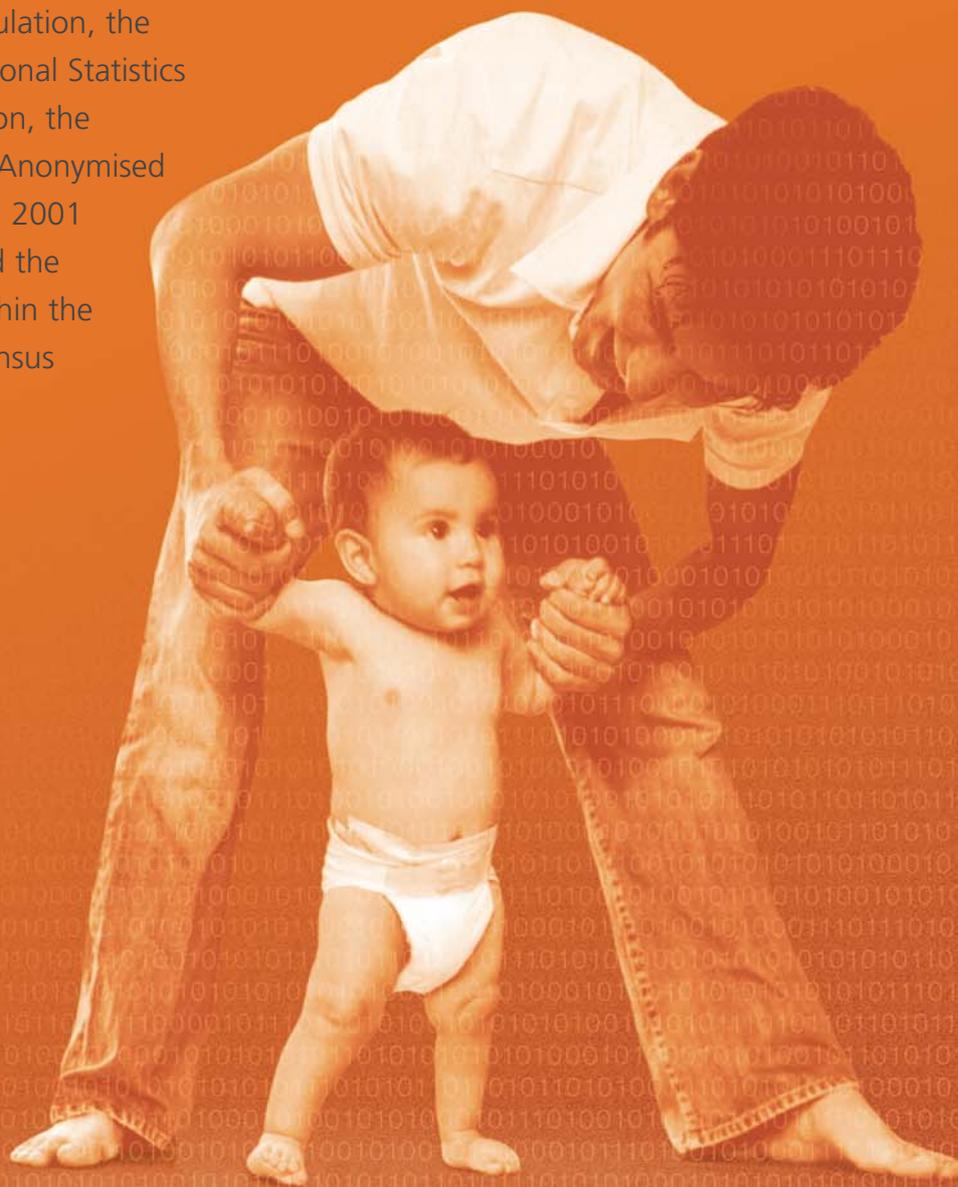
Engaging with plans for the development of population census and survey data

- 5.2 Plans for the 2011 Census of Population are already well advanced by the Census departments. This may be the last census involving the direct collection of information from households. Through the ESRC/JISC funded Census Programme, efforts will be made to ensure that these plans and new developments will meet the needs of the research community.
- 5.3 The ONS is also consulting with users over its plans for an Integrated Household/Continuous Population Survey which, from January 2008, will combine the Labour Force Survey, the General Household Survey, the Annual Population Survey, the Expenditure and Food Survey and the National Statistics Omnibus Survey within a single sampling frame. A similar approach is planned in Scotland. There is a need for the academic community to engage with the significant changes that will arise and to consider impacts on the continuity of research based on these sources.
- 5.4 The ESRC has already approved funding to continue the development of 'added-value' census resources through to 2011. Plans for funding beyond this horizon are dependent upon the shape of the Census 2021. With these plans in mind, the following action is proposed:

- **Representatives from the UK Census Departments and the ESRC will prepare a strategic plan to clarify the nature of their long-term collaboration over future population censuses. This plan will cover data access arrangements, historical continuity with existing census and survey data and the nature of future output requirements, particularly the need for detailed geographic information on population dynamics.**

In the development of national statistics

The Office for National Statistics and the ESRC have, for a number of years engaged in collaborations which have facilitated academic engagement with the important work of the ONS in developing national statistics. Examples include statistical work on the design of the 1991 and 2001 Censuses of Population, the development of a new National Statistics Socio-economic Classification, the creation of the Samples of Anonymised Records from the 1991 and 2001 Censuses of Population and the continued engagement within the framework of the ESRC Census Development Programme.



The development of administrative data resources

13 <

- 5.5 Much important information about individuals and organisations is retained nowadays through the use of information technology. The capture and storage of the details of transactions between organisations and individuals has created resources with the potential to inform social scientific research directly or, via data linkage techniques, to enhance existing resources. Such data cover a wide variety of fields in both the public sector and private sectors, including demographics, consumer behaviour, education, social care and community support, crime, transport, health, taxation, social security, housing and migration. Via personal or organisational identifiers, data from different sources can be linked to generate rich resources for research purposes. However, the scope for data linkage may be restricted by ethical considerations surrounding the sensitive nature of the linked data, by legal requirements and by the resource costs of undertaking linkage. There appears to be no consistent mechanism for striking a balance between ethical considerations and the public benefit from data linkage.
- 5.6 There is a need to establish the scope, quality and potential that public and private administrative data can offer as research resources. Given the wide variety of administrative data, work in this area needs to be focused on some important examples. Particular areas where attention could be directed include the registration and record system within the National Health Service, the Department of Work and Pensions Longitudinal Study of Benefit and Pension records, and the Department for Education and Skills (DfES) Pupil Level Annual Schools Census.
- 5.7 To progress work in this area, the following action is proposed:

- An audit of a selected number of major administrative databases held by government departments and agencies will be prepared. This audit will report on the potential such data have to inform research on a range of issues (with particular emphasis on data quality and coverage), the accessibility of such data for different research purposes (data anonymisation, the need for 'safe-settings', etc.), the feasibility of linking between different administrative data sources and with census and survey records. The audit will exemplify best practice linking already achieved or recently undertaken. This audit will be augmented by a study of public attitudes to data linking and the use of administrative data for research purposes. Work in this area will be undertaken as a close collaboration between the ESRC, relevant government departments/agencies and the ONS.

Reviewing longitudinal studies

- 5.8 Since 1991 the British Household Panel Study has been tracking the lives of 5,000 households across the UK. While there have been some significant enhancements to the size of the study over the past decade, attrition has now reduced the effective sample size. The panel must be renewed if such data are to continue to inform issues relating to demographic change and migration. A critical aspect of this renewal relates to the size of the new/enhanced panel study.
- 5.9 There is also a need to consider the long term future of the birth cohort and age cohort studies in the UK. Some of these (the 1958, 1970 and 2000-01 birth cohorts) have now been placed on firmer long-term funding than was previously the case. Other studies are funded on a periodic renewal basis.

› 14 5.10 To address these issues, the following actions are proposed:

- The ESRC has commissioned a review to establish the research requirements for a renewed panel study and to consider the future development of the birth and age cohort studies. The ESRC will consult with funding partners following this review and will act quickly upon the main recommendations arising from the review.
- The funding bodies for these resources (including the ESRC, MRC, Research Councils UK, the Office of Science and Innovation (OSI) and the DfES) will continue to work together as future plans for longitudinal studies are developed. This work will co-ordinate and link with those bodies with specific council or departmental responsibilities (eg through the Health-Related Longitudinal Studies Oversight Group, the ESRC Longitudinal Strategy Committee and the cross-departmental Longitudinal Studies co-ordinating Group).

Linking bio-medical and socio-economic data

5.11 A number of major data sources are now being developed which offer the potential to engage in research which addresses issues relating social and economic life to health and the development of medical conditions. Such data may combine information on genetics and other bio-medical details, medical records (including hospitalisation), mortality with social survey, census or other data from administrative sources. The potential to explore in detail the relationships between child development, diet, lifestyle, education, work, health and longevity that such linked data afford is vast. Data of this type may be subject to different procedures controlling their use for particular research purposes. There is a need to promote wider knowledge across the medical and social research communities about the potential such data offer for future research at this interface.

5.12 The following actions will be undertaken in 2006-07:

- The ESRC, MRC and the Wellcome Trust will continue to work together, and with other stakeholders, to develop and promote access to key data sources containing linked biomedical and socio-economic information.
- Plans will be developed by the ESRC, MRC and the Wellcome Trust to highlight the potential that linked socio-economic/biomedical data resources have to inform important research issues.

Enhancing research access to sensitive data

5.13 Sensitive data are defined as those which are potentially 'disclosive' (ie they could, in combination with other data, reveal the identity of an individual or organisation) or are protected under legislation which limits their distribution to researchers. Access arrangements to such data are varied. Some data collections are only available 'on site' – that is, in the premises of the data collection agency – and in a strictly controlled environment. Others are available only under special conditions regarding the way they are stored, analysed and destroyed after use.

5.14 Various procedures are currently being trialled to facilitate better access for research purposes whilst safeguarding the confidential nature of certain datasets. These range from new data licensing arrangements

to the development of secure virtual data laboratories. There is a need to review the problems of data access and the nature of data access arrangements, with the intention of improving and standardising access for research purposes.

5.15 To progress these developments the following action will be taken:

- **The ONS and the ESRC will undertake a joint review of recent developments in arrangements for research access to sensitive data. The review will seek to recommend ways in which access can be enhanced whilst simultaneously providing the necessary legal and ethical safeguards against misuse of such data.**

Developing access to business and management data

5.16 Private sector organisations hold significant amounts of personal information for a variety of reasons – either to improve customer services, to target markets for the goods and services they provide or in the interests of economic efficiency. Public sector bodies hold much information about organisations which is commercially sensitive (eg data supplied by commercial organisations to HM Revenue and Customs and to the ONS). Both types of data have value for research purposes, yet are seldom made available due to commercial sensitivity.

5.17 To explore the value and availability of business data for research purposes, the following actions will be undertaken:

- **The ESRC will organise and hold discussions with representatives of those holding such data to evaluate their potential as research resources and to determine how they might be made available for future research. Two reviews will be initiated: one focusing on the availability of commercial data for research purposes and one on the availability of organisation-based data (workplace surveys, management practices, productivity, innovation, etc.).**

Establishing better data resources for international and comparative research

5.18 To address issues relating to the changing global economy and migration and to facilitate a wider range of international comparative research, better data on individuals and organisations are required for a wider range of countries than is currently available. This will be achieved via negotiations for access with the appropriate bodies. To inform such negotiations and to promote international and comparative research on these important issues, the following actions are being pursued:

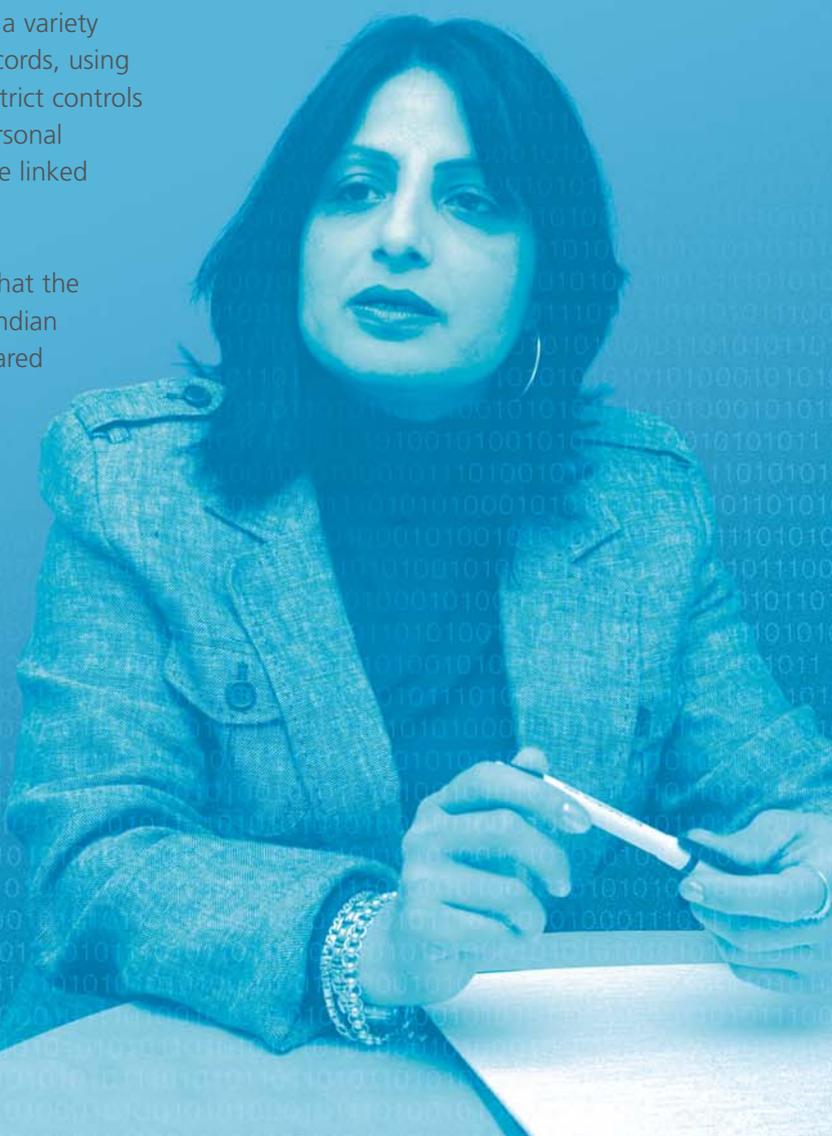
- **The ESRC has commissioned a review of international data resources and needs. This review will help shape a plan for the prioritisation of new data resources for comparative research and the improvement of access arrangements for the research community – a plan to be delivered in 2006-07.**
- **Building upon the strategic joint funding scheme announced by the ESRC and the Department for International Development, further work in this area will be undertaken in 2006 and involving a wider range of agencies (UK, EU and international), to explore the future international research agenda and to identify relevant data needs.**

Research on ethnic minority health in Scotland

Previous research has shown that there are substantial ethnic inequalities in health but there is little information on this in Scotland. A research project at Edinburgh University, which reported in 2005, used a linkage between Census information and records of illness and death to discover the true picture of health among minority ethnic communities.

Scotland has high-quality databases on hospital admissions. But they do not include full and reliable information about the patient's ethnic group. So the research project combined information from a variety of sources, including hospital and Census records, using computerised linkage. The project included strict controls to ensure the confidentiality of individual personal information, including restricted access to the linked database used for the research.

The research showed, among other things, that the incidence of heart attack amongst Scots of Indian and Pakistani origin is 60-70% higher compared to non-South Asians. The research helps provide the NHS in Scotland with the information and support it needs to ensure that Scotland's ethnic minority communities benefit fully from policies to address inequality and close the health gap.



6.0 MAKING THE NATIONAL DATA STRATEGY WORK

6.1 The work required to implement, carry out and maintain this strategic approach to the development of data resources for the social sciences is, in itself, resource-consuming. Recognising this, the ESRC, in collaboration with the MRC, the Wellcome Trust, the Nuffield Foundation, major UK government departments and the devolved administrations, have taken the following steps:

- Forming and participating in the **UK Data Forum**, a UK-wide committee consisting of representatives of the major government departments, research councils, the devolved administrations and other organisations involved in funding and developing data collection activities.
- Funding specific scoping studies and reviews to inform the future development of data resources.
- Funding the post of Strategic Advisor for Data Resources for an initial three-year period from October 2004 to September 2007, to help in the identification and the co-ordination of activities undertaken by government departments/agencies and the research councils which relate to the development of data resources.
- Drawing up the ESRC International Strategy in order to position the development of data resources in the context of global research interests and opportunities.

A number of key principles underlie this collaborative approach to developing, implementing and maintaining the strategy. These include:

Flexibility

6.2 Issues presently regarded as future research priorities may change or evolve as time progresses. For this reason the National Data Strategy will need to be updated periodically, linking evolving research issues to future data needs, identifying gaps in data provision and co-ordinating plans to meet such needs. The UK Data Forum is charged with this responsibility.

Establishing shared responsibilities

6.3 Where the development of a particular data resource may be deemed beneficial to a range of users; governmental and non-governmental, academic and non-academic, the UK Data Forum will work to support procedures through which these interests can be identified and responsibilities for development shared.

Prioritising future developments

6.4 Recognising the high cost of developing, establishing and maintaining some data resources, especially those which are longitudinal in nature, funding bodies may have to prioritise competing demands upon their limited resources. Mechanisms must be established through which competing claims to develop and maintain resources can be resolved according to agreed criteria.

Resolving data access issues

- 6.5 An important part of the research process is the nature of the arrangements which exist to regulate access to confidential or potentially disclosive data. High quality research requires facilities for good data access. This may involve linking between various data sources (eg linking administrative data to survey information held on individuals). Such procedures raise ethical issues relating to the confidentiality of data and the nature of consent required for linking data. Restrictions on access may also exist for legal reasons or because of the nature of undertakings to data providers given or implied during data collection. Funding agencies, data providers and data users will work together to identify these restrictions and facilitate the provision of guidance, development of best practice and common access arrangements whilst ensuring data are used in accordance with any necessary restrictions on use.

Promoting common standards for data access, description and documentation

- 6.6 While considerable advances have been made recently in standards of data preparation and documentation, not all data producers can or do provide data to common formats and with sufficient documentation to facilitate research. There is a need to co-ordinate efforts to ensure that common standards are adopted where appropriate.

Working across the countries of the UK

- 6.7 As the countries of the UK diversify further in terms of social, economic and health related policies, so they provide useful variation for scientific investigation of the links between different policies and their outcomes. This, in turn, stimulates the demand for harmonised statistical information relating to evaluation of these policy differences. Responsibility for the production of certain data resources now lies with the devolved administrations of Scotland, Northern Ireland and Wales. Although much work is currently undertaken or planned to ensure that the ONS and its counterparts in Scotland, Northern Ireland and Wales create UK-harmonised statistics, there are instances where the scale of data resources may be insufficient for certain statistical purpose in the countries of the UK. There is a need to ensure that efforts to produce 'UK-wide' data via the devolved administrations and the ONS are co-ordinated. The UK Data Forum will continue to monitor progress on this issue.

Linking with other strategic plans

- 6.8 The National Data Strategy links with and relates to other strategic plans already in place or under development across the countries of the UK. In particular, it recognises and builds upon a number of cross-departmental initiatives to identify future challenges for public policy (see Appendix). It must also be informed by and mesh with plans for the development and sharing of relevant data resources made by related UK bodies such as the MRC, Wellcome Trust and with similar developments across the European Union (eg the European Strategy Forum for Research Infrastructures).

Joint consultation over research priorities

19 <

- 6.9 The two main groups associated with the development of data resources – the research councils and government departments – adopt a variety of approaches to establish the research agenda which inform data needs and data development strategies. Government departments tend to equate the need for research evidence with policy – related requirements, both short term and long run while the Research Councils fund data collection and development activities to support longer-term research interests which may have no immediate policy-related implications.
- 6.10 There is, however, no clear cut distinction between what might be regarded as short and medium term policy-related needs and long term social and economic research. A review of recent consultations held by the research councils and government departments reveals a high degree of overlap in the nature of the research issues they identify and prioritise. Both groups would clearly benefit by sharing the thinking that informs their strategic planning and relating this to the need for data. An important element of the National Data Strategy will be to bring together all of the groups involved in the identification of research priorities in the social and economic sphere to meet formally at least once every two years. These Joint Consultations will bring together those with policy interests and social, medical and environmental scientists, to share their thinking regarding future research issues and priorities and to identify potential gaps in the provision of relevant data resources.

Data development, cost sharing and data sharing

- 6.11 When there is a strong shared interest in the development of new or existing data resources between the research councils, government departments and other funding bodies, mechanisms are required to translate this balance of interests into an agreed plan specific to the development of particular data resources. It is proposed that Data Development Agreements could be established in these instances. These agreements would be specific to a particular area of data development and would stipulate:
- The nature and funding of preliminary work required to scope a new data resource
 - The sharing of data development costs
 - Where appropriate, the sharing of the costs of maintaining a particular data resource
 - Arrangements for research access
 - Arrangements for sharing data resources and research outputs based upon these resources
 - Arrangements for documenting and archiving data resources.

Setting targets and reviewing progress

- 6.12 Targets and associated milestones provide the means through which desired progress can be measured, facilitating future revision of the plan or causing a refocusing of efforts. The National Data Strategy includes annual targets and milestones against which progress on its implementation can be gauged. The ESRC Strategic Advisor for Data Resources will regularly report progress on the implementation of the National Data Strategy to the UK Data Forum. Progress will also be reviewed periodically by the Research Resources Board of the ESRC.

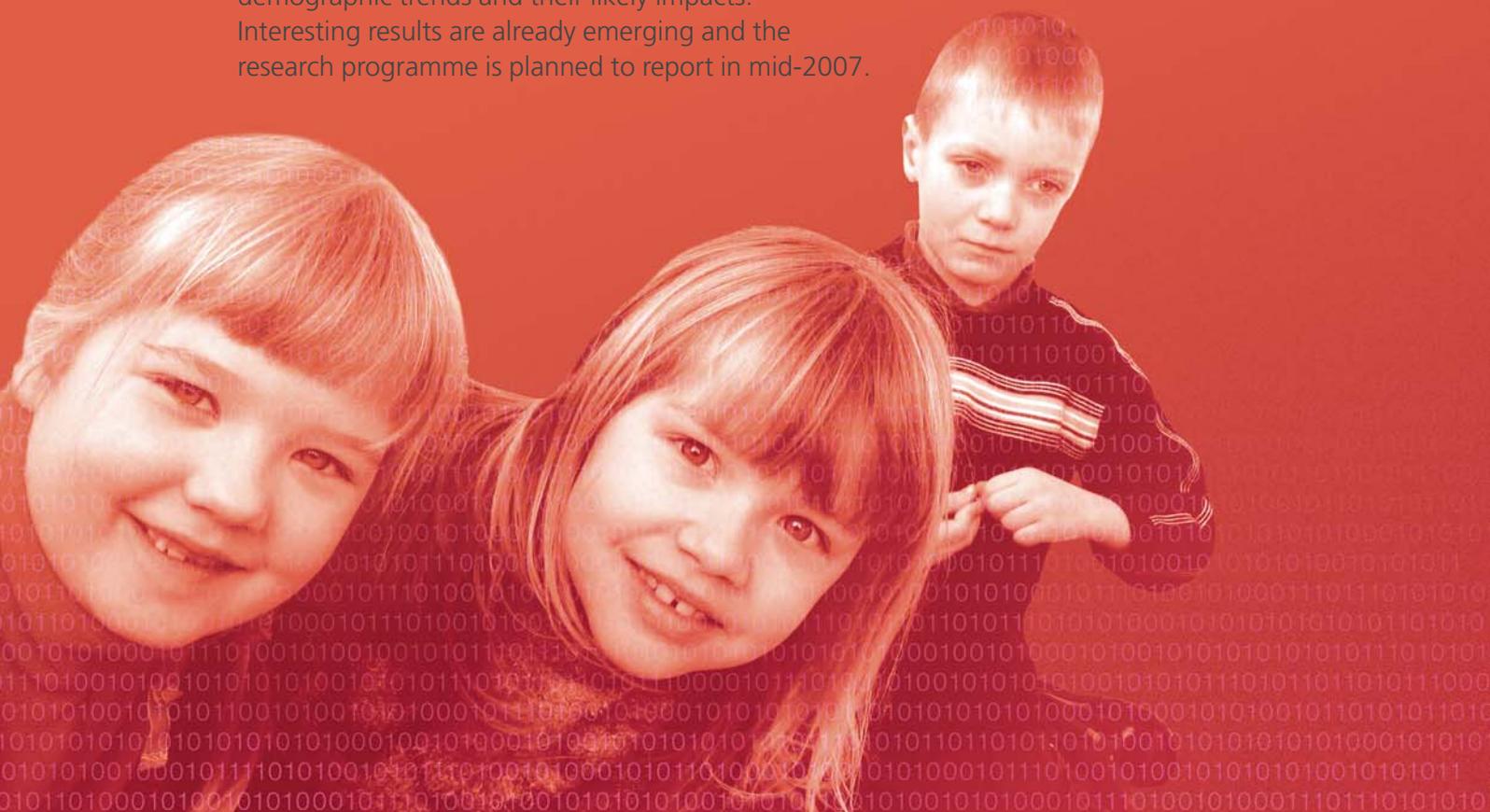
ESRC AND THE SCOTTISH EXECUTIVE

› 20

ESRC/Scottish Executive programme of research on Scottish demography

Scotland's recent demographic history is different from the rest of the UK. In particular, birth rates are relatively low and total population has slowly declined since the mid-70s (although with an upswing in the last 2 years). This poses important policy questions, in which there is much public and Ministerial interest. Yet relatively little is known about why some of the changes are occurring and therefore what policy measures might be taken.

In collaboration with the Registrar General for Scotland, the ESRC and Scottish Executive launched in early 2005 a 2 year £300,000 joint research programme which, following a call for proposals, funded 6 projects – 2 concerned with fertility, 2 with ageing and 2 with migration. The aim is to use existing sources of data to answer policy-relevant questions about the reasons for Scotland's different demographic trends and their likely impacts. Interesting results are already emerging and the research programme is planned to report in mid-2007.



Appendix

Initiatives identifying key research challenges

An important aim of the National Data Strategy is to identify data needs; to compare these needs with existing data provision, distinguishing gaps in the provision and to develop plans to address such gaps.

Future data needs are driven by the nature of the research challenges and issues that the scientific community will be addressing some 5 to 10 years from now. A number of initiatives have been developed to promote discussion and debate about the nature of these challenges. This annex gives a brief overview of the most significant of these initiatives and lists the issues that have been identified as future research priorities.

The ten-year Science and Innovation Framework

In July 2004 the government published this framework document². The report set out a strategy, supported by measurable targets towards progress, to achieve a number of goals designed to enhance the science base in the UK. These include the development of greater responsiveness of the publicly-funded research base to the needs of the economy and public services and the need to retain and build sufficient world class centres of research excellence.

The strategy is monitored on an annual basis and scrutinised by the UK Science Forum, a body of senior scientists, government advisors and civil servants³. The first monitoring report, published in July 2005⁴, details the measures being taken to improve the science base through the development of research infrastructure.

Alongside the Budget 2005, the government published *Long-term global economic challenges and opportunities in Europe*.⁵ Six key global economic challenges for Europe were identified including the need to take advantage of the increasing rewards from innovation as both global competition and the speed of technological change increase.

Council for Science and Technology

The Council for Science and Technology (CST), the Government's top level advisory body on science and technology policy issues, has examined a number of challenges associated with achieving the Government's vision for science and innovation. The CST is currently focusing on the Government's use of personal data⁶ to facilitate research; and on the use of scientific evidence about health in policy making across Government more widely.

Office of Science and Innovation (formerly OST) Foresight Directorate

The Foresight programme⁷ started in 1993. Foresight is the Government's science-based think tank, which provides evidence to Government on strategic cross-departmental issues with particular relevance to future policy. The current round, launched in April 2002 has adopted 'a new, more fluid, rolling programme of objectives', with a list of key challenges including research on obesity, intelligent infrastructure systems, the detection and identification of infectious diseases, brain science, addiction and drugs.

² http://www.hm-treasury.gov.uk/spending_review/spend_sr04/associated_documents/spending_sr04_science.cfm

³ Following a recommendation of the 2003 Lambert Review, the Government announced the creation of a high-level forum between Government, business leaders and scientists to support the UK's R&D and innovation goals and contribute to evidence informing future spending decisions. The Science Forum held its inaugural meeting on 20 July 2005.

⁴ <http://www.ost.gov.uk/policy/sifreview05.pdf>

⁵ http://www.hm-treasury.gov.uk/media/A1D/6B/global_final_140305.pdf

⁶ <http://www2.cst.gov.uk/cst/reports/#10>

⁷ 'Personal data' are data on individuals arising from administrative systems.

⁷ <http://www.foresight.gov.uk/index.html>

The Horizon Scanning Centre

In its *Science and Innovation Investment Framework 2004-2014*, the Government committed to establishing a Centre of Excellence in Horizon Scanning⁸, to be based in the Foresight directorate of the OSI. Work on establishing the Centre started in November 2004. Its output is intended to feed directly into cross-government priority-setting and strategy formation. The Centre describes its work as being 'strongly informed by the science base and by the best of existing work in Government, the private sector and elsewhere'. The Centre's aims are:

- To inform departmental and cross-departmental decision-making
- To support horizon scanning carried out by others inside and outside government
- To spot the implications of emerging science and technology and enable others to act on them.

The ESRC Research Consultation process

The ESRC undertook a major consultation⁹ about future research challenges and priorities in the spring and summer of 2004. This involved universities, learned societies, government departments, organisations in the business, public and voluntary sectors, and other key stakeholders. Among other recommendations, particularly about the global nature of research problems and research activity, the responses from the social science community identified the following research opportunities and challenges:

- Succeeding in the global economy
- Individual behaviour and its relationship to biological and social determinants
- Education for life
- Environmental change
- Security and international relations
- Religion, ethnicity and society
- Population change.

Co-ordination of research and analysis group

The Co-ordination of Research and Analysis Group (CRAG) was established in December 2004 'to promote more responsive analysis and dialogue between policy experts and the full range of analytical disciplines across government'¹⁰. Part of its remit is to identify analytical priorities and to promote a co-ordinated approach to analysis, especially where these priorities cut across the boundaries of a number of government departments and agencies. The CRAG organised a cross departmental seminar in June 2005 to prioritise analytical themes from a shortlist of topics. The following four were selected as potential priorities:

- Ageing population
- Migration
- Changing global economy
- Public service and service delivery.

The OSI Grand Challenges Exercise

The Department of Trade and Industry Five Year Strategy set the task of defining the 'grand challenges' facing public policy where research will play a major part in establishing policy directions. The Chief Scientific Advisers Committee met in March 2005 to discuss their input into this process. Two potential grand challenges have been identified. These are:

- Energy requirements and carbon emissions
- Coping with the changing profile of the UK population in terms of age, location and skills.

A third 'grand challenge' is under consideration, examining the need for better and more secure management of data.

⁸ http://www.foresight.gov.uk/HORIZON_SCANNING_CENTRE/index.html

⁹ http://www.lse.ac.uk/collections/researchAndProjectDevelopmentDivision/pdf/ESRC_stakeholder_consultation.pdf

¹⁰ http://www.gsr.gov.uk/gsr_network/crag.asp

ANNEX

Annex

UK Data Forum members

Chair
Professor David Rhind Vice-Chancellor, City University

Mr Richard Bartholomew Department for Education and Skills
Professor Colin Blakemore Medical Research Council
Mr Michael Bright Economic and Social Research Council
Dr Norman Caven Registrar General for Northern Ireland
Dr Kate Chamberlain Welsh Assembly Government
Mr George Clark Department for Work and Pensions
Dr Alan Doyle Wellcome Trust
Mr Keith Dugmore Demographic Decisions Ltd
Ms Sue Duncan Cabinet Office
Ms Karen Dunnell Office for National Statistics
Mr Nick Dyson Her Majesty's Revenue and Customs
Professor Peter Elias Economic and Social Research Council,
Strategic Advisor (Data Resources)

Mr David Frazer Department for Work and Pensions
Ms Sharon Jones Department for Work and Pensions
Mr Duncan Macniven General Register Office for Scotland
Dr Jil Matheson Office for National Statistics
Mr Jeremy Neathey Economic and Social Research Council
Dr Malcolm Read Joint Information Systems Committee
Dr Bridget Rosewell Office of the Mayor of London
Dr Caroline Stone Medical Research Council
Professor Paul Wiles Home Office
Dr Sharon Witherspoon Nuffield Foundation

“Longitudinal studies and the collection and maintenance of national datasets are essential to build an evidence base to inform effective policy development. Such work is, however, costly and it is unreasonable to expect ESRC to fund these major activities out of its modest budget. Where a dataset is of particular relevance to a Government Department, the Department should shoulder the majority of the financial burden, whilst taking advantage of ESRC’s skills and experience. We recommend that ESRC pursues this issue with the relevant Departments.”

HOUSE OF COMMONS SCIENCE AND TECHNOLOGY COMMITTEE
THE WORK OF THE ECONOMIC AND SOCIAL RESEARCH COUNCIL
FIRST REPORT OF SESSION 2004-05





The *National Strategy for Data Resources for the Social Sciences* is a plan which aims to ensure that the national data infrastructure meets the demands which will be placed upon it to address both current and future research needs in the social sciences. It has been developed by The UK Data Forum, a body which brings together those bodies with responsibilities for the development of data resources across two research councils, research foundations, a range of government departments, the Office for National Statistics and the devolved administrations.

The UK Data Forum is serviced by the ESRC Strategic Advisor for Data Resources. If you have any comments on this document, or if you want to contact the UK Data Forum about other issues you consider relevant to the National Data Strategy, please contact:

Professor Peter Elias
ESRC Strategic Advisor for Data Resources
Institute for Employment Research
University of Warwick
Coventry CV4 7AL

Tel: +44 (0)24 7652 3286
Fax: +44 (0)24 7652 4241
Email: peter.elias@warwick.ac.uk



The Economic and Social Research Council is the UK's leading research and training agency addressing economic and social concerns. It aims to provide high-quality research on issues of importance to business, the public sector and Government. The issues considered include economic competitiveness, the effectiveness of public services and policy, and our quality of life.