Statistical Potential of Administrative Records

An Examination of Data Holdings in Six Government Departments

Working Report

For further information, please contact Gerry Brady, Social Statistics Integration Division, CSO, 01 498 4006

Central Statistics Office

September 2003

Contents

Tables

Abbreviations

BCG	Bacillus Calmette-Guerin vaccine against Tuberculosis
BUPA	British United Provident Association
CAO	Central Applications Office
CARS	Classifications and Related Standards
CAT	Capital Acquisitions Tax
CB	Child Benefit
CCEI	Central Client Eligibility Index
CCTS	Courts Case Tracking System
CIDR	Computerised Infectious Disease Reporting
CIS	Client Identity Services
CN	Combined Nomenclature
COP	Census of Population
CPA	Classification of Products by Activity
CPC	Central Product Classification
CRO	Companies Registration Office
CRS	Civil Registration Services
CRS	Central Records System
CSO	Central Statistics Office
DB	Disability Benefit
DES	Department of Education and Science
DETE	Department of Enterprise, Trade and Employment
DHC	Department of Health and Children
DIRT	Deposit Interest Retention Tax
DJELR	Department of Justice, Equality and Law Reform
DOB	Date of Birth
DoEHLG	Department of the Environment, Heritage and Local Government
DPS	Drugs Payment Scheme
DSFA	Department of Social and Family Affairs
ED	Electoral Division
EEA	European Economic Area
EPA	Environmental Protection Agency
ERC	Education Research Centre
ERHA	Eastern Regional Health Authority
ESF	European Social Fund
ESI	Educational Services Interactive
ESRI	Economic and Social Research Institute
EUROCAT	European Concerted Action on Congenital Anomalies and Twins
EU-SILC	EU Statistics on Income and Living Conditions
FÁS	Foras Áiseanna Saothair
FOI	Freedom of Information Act
GCEB	Giving Children an Even Break
GIS	Geographic Information Systems
GMS	General Medical Services
GMSPB	General Medical Services (Payments) Board
GNIB	Garda National Immigration Bureau
GRO	General Register Office
HBS	Household Budget Survey
HBSC	Health Behaviours in School-aged Children
HEA	Higher Education Authority
HIPE	Hospital In-patient Enquiry
HRB	Health Research Board
HS	Harmonised System
IALS	International Adult Literacy Survey
IBC	Irish Born Child
ICBD	International Clearinghouse for Birth Defects Monitoring Systems
ICD	International Classification of Diseases
ILO	International Labour Organisation
INSPIRE	Infrastructure for Spatial Information In Europe

ISCED	International Standard Classification of Education
ISDI	Irish Spatial Data Infrastructure
ISIC	International Standard Industrial Classification
ISO	International Standards Organisation
ISTS	Integrated Short Term Schemes
JCSP	Junior Certificate School Programme
LA	Local Authority
LCVP	Leaving Certificate Vocational Programme
LGCSB	Local Government Computer Services Board
NACE	General industrial classification of economic activities within the EU
NAPS	National Anti-Poverty Strategy
NBSB	National Breast Screening Board
NBSP	National Breast Screening Programme
NCEA	National Council for Educational Awards
NCRB	National Cancer Registry Board
NCRI	National Cancer Registry of Ireland
NCS	National Children's Strategy
NCT	National Car Test
NCTE	National Centre for Technology in Education
NDP	National Development Plan
NDSC	National Disease Surveillance Centre
NEPS	National Educational Psychological Service Agency
NESC	National Economic and Social Council
NESF	National Economic and Social Forum
NHIS	National Health Information Strategy
NIDD	Notifiable Infectious Diseases Database
NISRA	Northern Ireland Statistics and Research Agency
NPDF	National Planning Data Form
NPIPRS	National Psychiatric In-Patient Reporting System
NSB	National Statistics Board
NSHQ	National Survey of Housing Quality
NSS	National Spatial Strategy
NUI	National University of Ireland
NUTS	Nomenclature des Unites Territorial Statistique
NVDF	National Vehicle and Driver File
ODTR	Office of the Director of Telecommunications Regulation
OECD	Organisation for Economic Co-operation and Development
OIB	Occupational Injury Benefit
OLR	On-Line Review
ONS	Office for National Statistics (UK)
OSI	Office for Social Inclusion
PAYE	Pay As You Earn
PD	Particulars Delivered
PENLIVE	Pensions Live system
PES	Principal Economic Status
PISA	Program for International Student Assessment
PMI	Precision Marketing Information Ltd
PPD	Primary Pupils Database
PPF	Programme for Prosperity and Fairness
PPS Number	Personal Public Service Number
PQ	Parliamentary Question
PRETA	Pre-Retirement Allowance
PRIS	Prisoner Records Information System
PRSI	Pay Related Social Insurance
PRTB	Private Residential Tenancies Board
PSI	Public Service Identity database
PULSE	Police using Leading Systems Effectively
QNHS	Quarterly National Household Survey
RAC	Refugee Applications Commissioner
RACTS	Registration and Case Tracking System

RAT	Refugee Applications Tribunal
Revenue	Office of the Revenue Commissioners
RIA	Reception Integration Agency
SDI	Spatial Data Infrastructure
SDM	Service Delivery Model
SES	Socio-Economic Status
SGSES	Steering Group on Social and Equality Statistics
SLAN	Survey of Lifestyles, Attitudes and Nutrition
SnoMed	Standard Nomenclature for clinical terms
SOGSI	Senior Officials Group on Social Inclusion
SP	Service Provider
SPAR	Statistical Potential of Administrative Records
SPC	Social Protection Committee
SW	Social Welfare
SWA	Supplementary Welfare Allowance
TCA	Tax Consolidation Act
UA	Unemployment Allowance
UB	Unemployment Benefit
UNESCO	United Nations Educational, Scientific and Cultural Organisation
UPI	Unique Patient Identifier
VHI	Voluntary Health Insurance
VPT	Vocational Preparation Task
VRT	Vehicle Registration Tax
VRU	Vehicle Registration Unit
VTOS	Vocational Training Opportunity Scheme
WHO	World Health Organisation

Preface

Arising from the report of the National Statistics Board (NSB) on social and equality statistics[1], the Government requested the CSO to take a lead role in the development of the potential of administrative data across government departments and agencies in conjunction with the Senior Officials Group on Social Inclusion (SOGSI). The Government also decided that a formal data/statistics strategy be developed within each department as part of its information strategy; and that the NSB, supported by the CSO, should develop best practice guidelines for departments for the preparation and implementation of a formal data/statistics strategy.

In response to the Government request, the CSO established teams to examine the statistical potential in six government departments and to prepare a report by end-July. The examination was designed to focus mainly on areas of social statistics and also to identify the potential additional value to be gained if some of the administrative data holdings could be integrated.

This report should be considered to be a "working report" and reflects the interactions between the CSO teams and a selection of contacts in the six departments. Nevertheless, because some key issues are addressed and valuable recommendations made, the report is being made widely available and has also been released on the CSO website (www.cso.ie). The intention is to continue the review of the potential for social statistics in a further three or four departments; and then to consider (with the NSB) the potential of administrative data for other fields of statistics.

Chapters 2 to 7 of this report present the analyses and main recommendations of the teams in each of the six departments. Chapter 8 addresses some key issues which would facilitate the integration of data for statistical and analysis purposes across government departments and which would ultimately be of benefit to the public in the form of better policy responses and more cost-effective public administration. Among these important issues are the wider use of the Personal Public Service (PPS) Number; the benefits and efficiencies of using standard classifications and coding systems as widely as possible; the importance of structuring data to support the National Spatial Strategy and the provision of regional data generally. The report concludes in chapter 9 with a list of 21 recommendations; nine of which are specific to particular departments and twelve of which are of a cross-cutting nature to facilitate the development of a more comprehensive and integrated system of statistics across the public service. The recommendations are best understood in the context of the discussion in the various chapters.

Readers will be struck by the key, central importance of the PPS Number in integrating data for statistical and analysis purposes. It is not difficult to think of situations where a wider use of the PPS Number could be of direct benefit to individuals and to society. Therefore, as recommended by the NSB, the CSO will be developing protocols in relation to the proper use of administrative data for statistical purposes within a framework which is acceptable to the Data Protection Commissioner.

Donal Garvey
Director General

[1] NSB (2003): “Developing Irish Social and Equality Statistics to meet Policy Needs”

Chapter 1

Background

1.1 Introduction

This report presents the findings of an examination of the data holdings in six government departments[2]. The report is primarily focussed on data holdings in the six departments of relevance to social statistics. The report was compiled by the CSO in response to a Government decision that the CSO should take the lead role in the development of the potential of administrative data across government departments and agencies and other bodies reporting to these Departments[3]. This chapter describes the background to this decision and the initial work being undertaken by the CSO to implement it.

1.2 National Statistics Board Strategy 2003-2008

The NSB is responsible to Government for setting priorities for the compilation and development of official statistics in Ireland and for guiding the strategic direction of the Central Statistics Office (CSO). In its Strategy 2003-2008[4], the NSB outlined its approach for developing the Irish Statistical System. It recommended a networked approach, across government departments and agencies, to ensure that statistics were delivered cost-effectively and that statistical surveys and information collected through administrative sources were fully integrated into the system of official statistics. Such a whole-system approach was seen as necessary to ensure that Ireland could produce statistics that permitted objective evaluation and monitoring of economic and social progress in Ireland. In the case of administrative data, the NSB recommended a strategy whereby:

- Data in individual departments and agencies would be capable of integration both within and across departments;
- Data in individual departments and agencies would be capable of being used for longitudinal analysis; and
- The burden on users of State services and those delivering the services in respect of generating data capable of statistical analysis would be minimised.

The NSB concluded that a network approach across government departments was required to implement this strategy effectively. The NSB proposed that the CSO should take a central role in the development of the potential of administrative data across government departments in conjunction with senior cross-department management groups, such as the Senior Officials Group on Social Inclusion (SOGSI) in the case of social statistics[5]. The NSB's view is that the provision of adequate statistical information should be regarded as a desired output of administrative data sources in the public sector. The statistical potential of these data sources should be assessed, the potential fully developed where appropriate, and greater consistency in their statistical management should add significantly to their information value.

[2] Education and Science; Environment, Heritage and Local Government; Health and Children; Justice, Equality and Law Reform; Social and Family Affairs; and the Office of the Revenue Commissioners.
[3] For convenience, these departments, agencies and bodies will generally be referred to as government departments throughout this report
[4] NSB (2003), "Strategy for Statistics 2003-2008"
[5] The central role for the CSO in the development of administrative records for statistical purposes is already established in the Statistics Act, 1993.

1.3 Steering Group on Social and Equality Statistics

A Steering Group on Social and Equality Statistics (SGSES) was established in mid-2002 to undertake a scoping study on Irish social and equality statistics. The group comprised key policy and statistical staff in relevant government departments, CSO staff and experts with experience in analysing social data in Ireland. The group published its report in April 2003[6]. The Government memorandum accompanying the SGSES report contained the following provisions:

- The CSO will take a lead role in the development of the potential of administrative data across government departments and agencies in conjunction with the SOGSI. In particular, two actions are essential to tap into the potential data resources at both departmental and agency level:
  - Increased standardisation, co-ordination and classification of data collection and maintenance to be agreed inter-departmentally; and
  - Investigation of the expanded use of common identifiers, such as the Personal Public Service Number (PPS Number) or postal codes, if developed.

The Government and social partners included this commitment in the social partnership agreement 2003-2005, *Sustaining Progress*[7].

1.4 Examination of data holdings in six government departments

With a view to giving effect to the commitments set out in the SGSES report, the Director General of the CSO established six CSO teams to examine the statistical potential of data holdings in six government departments. This initial examination, known as the SPAR project (Statistical Potential of Administrative Records), was designed to identify the potential value in the individual data holdings in these departments and the potential additional value to be gained if some of these data sources could be integrated. The membership of the teams comprised a mix of experienced Senior Statisticians and Statisticians in the Cork and Dublin offices of the CSO (see Appendix 1). The Director General secured the support of the Secretary General of each of the six departments and agreed contact persons in the departments for the work of the SPAR teams.

The following were the terms of reference given to the teams by the Director General:

- To prepare a report, by end-July, on the statistical potential for key policy purposes of departmental data holdings.
- The report should take account of the variables already available in certain data holdings and how they are captured (whether in standard or non-standard formats). It should comment on the additional value to be gained by capturing the data according to recognised statistical classifications and standards.
- In the context of cross cutting issues, the report should make proposals on the added value potential of capturing as part of the data sets a linking capability (the

[6] NSB (2003): "Developing Irish Social and Equality Statistics to meet Policy Needs"
[7] Department of the Taoiseach (2003): "Sustaining Progress, Social Partnership Agreement 2003-2005"

PPS Number) or additional or re-oriented variables (e.g. date of birth instead of age).

- While the focus of the SGSES work related to social statistics, this effort should take into account (to the extent feasible) other important departmental data holdings that may exist.

The main objective of the project was to identify possibilities for enhancing the long-term statistical potential of the data holdings in the six government departments.

Working procedures

Initial internal meetings were held, where the teams reviewed the SGSES report and agreed a questionnaire to ensure that a consistent approach was taken across the six departments. The teams were also briefed on the main points of the NSB's 2003-2008 Strategy to ensure that their approach was consistent with the principles outlined in the Strategy. The NSB proposed that all official statistics produced in Ireland should be compiled to best international standards and that all the cross cutting potential of existing data sources in government departments should be fully realised.

The teams conducted their work with the six departments, and their associated bodies, through a mixture of direct meetings, telephone and e-mail contact. As well as holding a considerable number of meetings directly with contact officials in each department, the teams also met with staff in a number of departmental bodies, e.g. the Higher Education Authority and the General Medical Services (Payment) Board. In addition, the teams presented their draft chapters for comment to the staff with whom they worked in the various departments. The project was co-ordinated through regular meetings of the six team leaders and support staff mainly from the Social Statistics Integration Division of the CSO.

The CSO, on behalf of the teams, would like to thank the staff in the six departments for their considerable input into this report.

1.5 Brief outline of the report

Chapters 2 to 7 present the findings of the teams in each department. The findings are generally presented as reviews of the individual major data sources or systems in the six departments.

Chapter 8 describes the Client Records System (CRS) operated by the DSFA and the related issue of the use of the PPS Number. Chapter 8 also outlines the key recommendations of the teams on common inter-departmental issues such as the benefit of using standard data classifications, statistical management of the individual data sources and other cross-cutting issues.

Chapter 9 summarises the key conclusions and recommendations of the teams.

Chapter 2

Department of Education and Science

2.1 Role and functions of Department

The mission of the Department of Education and Science, as set out in its statement of strategy 2003-2005, is to provide for high-quality education, which will:

- Enable individuals to achieve their full potential and to participate fully as members of society; and
- Contribute to Ireland's social, cultural and economic development.

This mission statement explicitly reflects the dual aim of the education system. It is focussed on the needs of the individual. It is also focussed on the needs of society. The Department is involved in the provision of education at pre-school, primary and post-primary school, and at universities and other third level institutions as well as adult and further education.

The major functions of the Department are in respect of policy formulation, resource allocation and the evaluation of performance and educational outputs. These three major functions are of course interlinked, as the evaluation of performance and outputs from the education system feed directly into the areas of policy formulation and resource allocation.

In recent decades, there has been a growing awareness that education is broader than just the delivery of curricula in schools. Education is seen as a lifelong process in which schools play a crucial role. Data on educational spending, resources, enrolment and outcomes need to be placed in the context of lifelong learning and the information needs of various stakeholders. The quality of educational outcomes and their relationship to learning processes in schools has come to the fore in international comparisons as well as policy analysis at national level. There is increased pressure to go beyond statistics of enrolment, teachers and spending - crucial as these are to policy planning and information - and to develop indicators of system performance and quality of educational outcomes, inputs and processes.

Some key areas of policy priority relate to:

- Social inclusion and broader educational equality and equity issues. The most recent partnership agreement[8] states that "education has a fundamental role in providing full access to life chances and breaking the cycle of disadvantage";
- Benchmarking of skills in the adult population;
- Indicators of educational achievement at primary and second level;
- Trends (past and projected) in enrolment, teaching staff and expenditure at each level of education;
- International comparisons across all areas of education;
- School-level indicators of performance;
- Special needs education and associated indicators and measures; and
- Participation in education, training and learning outside the formal education sector (including early childhood education and care). This particular area underscores the necessity for cross-departmental co-operation.

[8] Department of the Taoiseach (2003): “Sustaining Progress, Social Partnership Agreement 2003-2005”

Traditionally, the Department has been focussed on operational matters and educational inputs. However, following on from a report on the Department's operation, systems and staffing, chaired by Mr. Séan Cromien (former Secretary of the Department of Finance), the Department's latest Statement of Strategy has highlighted the "absolute need to shift from its overwhelming focus on day-to-day operational matters to a situation where serious attention could be given to the vital areas of policy development, planning and evaluation". This up-front acknowledgement that, to date, the Department has not been in a position to give serious attention to policy development and evaluation provides a useful insight into how departmental data holdings have been regarded historically. The Statement of Strategy commits the Department to developing a stronger planning capacity and to making better use of existing evidence and research.

Location of data holdings and role of the Statistics Section

While there is quite an amount of data in existence, meeting the key policy needs of the Department will require more than a greater appreciation of the existing data. There are areas where it will require more fundamental evaluation of the type of data that should be collected and held by the Department. This is because the range of sources currently available to the Department (surveys and administrative sources) allows identification of outcomes at a general level only. The current sources do not, however, provide an adequate basis for decisions on educational interventions and monitoring at the level of the individual pupil.

There are, in all, 27 bodies or agencies under the aegis of the Department. Some of these bodies maintain their own administrative data holdings. The Higher Education Authority (HEA) is perhaps the largest body falling under this category. The National Centre for Technology in Education (NCTE) also maintains its own data sources. Also, quite an amount of research is conducted by outside organisations - the Educational Research Centre (ERC), the ESRI and many other external organisations conduct research commissioned by the Department or by its agencies. Where research is conducted by an external organisation, the norm is that the raw micro data are held by that external organisation.

Within the Department, data holdings are held and maintained by line sections. This has led to a situation whereby data holdings are not well connected in a physical or conceptual sense.

Within the Department there is a small Statistics Section. This section consists of a Senior Statistician, a Statistician and seven (full-time equivalent) administrative staff. The bulk of the work of the administrative staff is taken up with the collection of information for the Annual Census of Primary Schools and the collation of aggregated data from other sources (other line sections within the Department and outside agencies such as HEA). Much of this work is for the purpose of producing the Department's Annual Statistical Report, which is a comprehensive publication giving detailed information primarily concerning educational inputs (number of pupils, number of teachers, certain expenditures). The Senior Statistician and Statistician service policy areas of the Department by collating and analysing data - for instance projecting future enrolments and supply and demand for teachers, sourcing data for PQ responses, reporting on equity issues at third level, interpretation of results of educational research, etc. Servicing international requests for information (from EU, OECD, UNESCO) has also been consuming a growing proportion of the resources of the section.

Approach of SPAR team

The initial approach adopted by the DES SPAR team was to list and prioritise the departmental data holdings with a view to investigating:

- The potential for aligning the data sources more closely with departmental needs;
- The potential to draw on existing information from data holdings of other departments; and
- The potential to share information from DES data holdings with other departments with a view to meeting their data needs.

In essence, the SPAR project was to be a "bottom up" exercise which primarily focussed on data sources. This approach, however, was predicated on having detailed information about the contents of each data source. The main point of contact with the Department was with the staff of the Statistics Section, who were not in a position to readily supply lists of variables together with coding frames. In some cases the documentation required did not exist in readily accessible formats. In other cases, it simply was not possible to contact all of the relevant outside agencies in the time available. There were, in all, 34 sources listed as a result of the SGSES exercise. With the benefit of hindsight, it is now apparent that the Department was probably too thorough in its compilation of the initial list. Line sections listed sources that were essentially once-off surveys rather than major or recurring data sources. However, even a superficial appraisal of these sources by the SPAR team and the staff of the Statistics Section was time consuming. It is worth noting that the lack of a structured way of sharing information (e.g. groupware) within the Department made it difficult for the staff of the Statistics Section to point us in the precise direction of all relevant information.

In the event, even a preliminary appraisal of the 34 listed sources led us to consider the coherence of the data holdings and the purposes they served. This led us back to consideration of departmental needs. This could be considered a "top-down" approach. The contents of this chapter are, therefore, the output of a process that attempts to marry the main needs to the supporting data infrastructure. This was a difficult and ambitious task as the development of a stronger statistical and information base in the Department will require a more precise articulation of the links between statistics, needs for policy purposes, administrative needs and evaluation than is available at present.

Departmental needs and sources

An inventory of data needs was supplied for the SGSES report. This work was also co-ordinated by the Statistics Section. All policy sections were contacted and each specified its own needs. As well as ongoing administrative needs, many newer needs (generated by NAPS, NDP and Sustaining Progress) were included. Also, each need was categorised as to whether or not it was being met and by a priority indicator. From this, it is apparent that many high-priority needs have yet to be met. In fact, it is difficult to see how some of the needs could be met from existing sources. Also, of the higher priority needs that are being met, there are question marks over the quality of some of the indicators in use[9].

The issues of educational disadvantage (see Section 2.5) and educational interventions are recurring features of the data needs of the Department. The importance of early identification of educational disadvantage and subsequent intervention is considered critical.

[9] This latter point is particularly relevant to those indicators that refer to, or need to be disaggregated by, socio-economic status or educational disadvantage

Disaggregation of many indicators by ethnicity, socio-economic status and school disadvantaged status are included in the list of needs. However, at primary level, no data at individual pupil level are held by the Department. Annual reporting is at the level of the school, grade and class. Making recommendations involving minor changes to this source would not be productive, since the Department has recognised that a Primary Pupil Database (at individual level) needs to be introduced. Instead, we propose a possible solution that may enable an elementary database to be operational relatively quickly, whether as an interim measure or otherwise.

At tertiary level, the Department is responsible for the administration of Institutes of Technology and certain other institutions. The HEA is responsible for Universities and some other colleges. At present, the type of data holdings differ according to the type of institution. However, it is planned that the HEA will assume responsibility for all third level institutions in 2004. In light of this, we spoke directly to the HEA to establish their plans for data holdings from 2004, rather than focussing on the existing third level sources maintained by the Department.

2.2 Thematic categorisation of needs and grouping of sources

Instead of addressing the stated needs individually, we deemed it more appropriate to categorise them thematically. This made it possible to deal with every source (or group of sources) and to consider the potential of each source to meet needs around each theme. This was done in recognition of the fact that the expressed needs may not be fixed - we would expect that improving the potential of a source would lead to a refocusing of needs. The themes we identified were:

- Equity and equality issues, including access, educational disadvantage, and educational interventions;
- Attainment and achievement, retention issues, and effectiveness of the educational system; and
- Life outcomes for individuals.

Obvious themes that are not included are expenditure, physical infrastructure and enrolment projections.

The data sources were categorised as follows -

- Primary level sources - the Annual Census of Primary Schools;
- Second level sources - the Post Primary Pupils database;
- Surveys of achievement at Primary and Secondary level (ERC);
- State Examination Results;
- Third level administrative sources, and the National Survey of Access to Higher Education;
- Destination of exits surveys - Survey of School Leavers (ESRI) and the First Destination Report (HEA) ; and
- General Population Surveys (QNHS, EU-SILC, HBS, IALS).

2.2.1 Measuring inputs and outcomes - potential use of PPS Number

It is sometimes assumed that data needs for administration are different to data needs for statistical purposes. However, it could be argued that this is not the case in DES. The mission statement makes clear that a primary focus of the education system is on the needs of the individual. Many of the needs expressed by the Department relate to educational outcomes (for the individual). Many of the newer needs reflect a focus on targeted outcomes for educationally disadvantaged children[10]. There are many schemes targeted at disadvantaged children (or more often, targeted at schools deemed to have measured concentrations of disadvantaged children). However, in order to target disadvantaged children, one must first identify them. In order to assess the effectiveness of the targeted educational interventions, one must have the capacity to measure the outcomes for those individuals. In other words, there needs to be a strong capacity for longitudinal tracking of individuals through the education system - and beyond. There is already an amount of information held system-wide which relates to certain outcomes for individuals through the social welfare system and revenue files, and these sources hold the PPS Number. However, currently there is no systematic method used to relate these particular individual outcomes to educational inputs - the initial demographic and socio-economic characteristics of the individual, the educational institutions attended, the types of educational interventions or the levels of attainment reached.

At the initial stage of the educational cycle (primary level) no information is currently held on the background of the individual pupil. However, DSFA holds much information relating to the demographic and socio-economic situation of families (through means tested schemes and the child benefit system).

A recurring feature of this chapter is the highlighting of the role that could be played by Revenue and DSFA sources in measuring both inputs to the educational system and outcomes for individuals through more systematic use of the PPS Number.

2.2.2 Geographic coding, catchment areas, and enrolment projections

Many departmental sources include address information. In many cases (pupil/student databases, whether existing or proposed) the addresses are those of individual students. Usually, the address data is coded to county level (or for Dublin, postal district). The availability, and use of, an automatic geo-coding system across all departments would greatly enhance the value of this data, while simultaneously reducing the overheads involved in assigning county codes to address strings.

For instance, there are reasons why the Department might be interested in the catchment area of schools or third level institutions. Also, when projecting enrolments, the Department's practice is to project from the number of births, while making assumptions about inward and outward flows. As PPS Numbers are assigned to every child born and the address of a child is held on the DSFA Central Records System (for the purposes of administering the child benefit scheme), it may be that the accuracy of enrolment projections at a local level could be enhanced.

In Section 8.3 there is a description of developments in unique addresses and geo-coding.

[10] The Education Act defines educational disadvantage as "the impediments to education arising from social or economic disadvantage which prevent students from deriving appropriate benefit from education in schools".

2.3 Data Sources

2.3.1 Primary level sources - the Annual Census of Primary Schools

The Annual Census of Primary Schools is conducted by means of paper questionnaire to almost 3,300 principals of aided primary schools. The returned questionnaires are keyed by staff of the Statistics Unit (for the most recent census, this work was outsourced). School principals complete the questionnaire. The information requested is at aggregate level. A row of the questionnaire is allocated to each class within the school. In each row, there are fields for number of pupils by gender, as well as number of pupils by year of birth. Some information (at school level) on the number of children with special needs taught in special classes within ordinary schools is included - the Department uses its own disability coding structure (13 types in all). Although called "disability", this variable includes codes referring to special or social needs as well as disability.

It is acknowledged by the Department that this source is inadequate for the needs (policy and administrative) of the Department. Central to this thinking are issues around the designation of disadvantage and the allocation of resources to counter disadvantage. The Department uses the concept of "disadvantaged school status" as a means of deciding on the type and level of additional resources that should be allocated to a school. This status is allocated partly by estimates of the numbers of pupils whose families qualify for medical cards. These estimates are made by school principals. It is not clear whether the estimation procedure is carried out using the same objective criteria by all principals. Many of the listed performance or effectiveness needs at primary level include disaggregation by "school disadvantaged status". Some also refer to disaggregation by ethnicity, refugee status, membership of the traveller community as well as socio-economic status.

A departmental priority is the establishment of a Primary Pupils Database - at individual pupil level. A project team within the Department is currently undertaking a requirement analysis. This is part of a wider initiative - the Educational Services Interactive Project (ESI). In Section 2.5, we outline an approach that the Department may wish to consider as an interim solution, pending the development of a more advanced system.

One of the major advantages of having a Primary Pupils Database would be that pupil PPS Numbers could be collected. It would be difficult to overstate the value of this information. In terms of educational disadvantage, it should facilitate designation of disadvantaged status to an individual rather than a school, by referencing data on means tested schemes in DSFA. In terms of disability and special needs, an indicator could be attached to the individual - again, using records of other departments that include the PPS Number. Obviously, not all relevant information may be available from other departmental holdings. The systematic collection of data on variables related to educational disadvantage and other variables known to affect the educational attainment/achievement of children needs to be seriously considered. There are issues around whether that data should be collected by schools and reported to the Department or whether the collection should be done centrally by the Department (for instance postal questionnaire to the homes of all entrants to the primary system). The latter would be appropriate if it were considered that certain relevant characteristics were sensitive and need not be known to the school or teacher for educational purposes.

As regards educational interventions or specific schemes addressing educational interventions, an indication of the intervention or scheme from which the pupil has benefited could be included in the record (rather than indications of schemes that the school has been entitled to). This would facilitate evaluation of the effectiveness of various schemes. For instance, when choosing a sample to test literacy or numeracy attainment at various stages of the cycle (see section 2.3.3), having access to demographic and socio-economic

information, as well as information on interventions at individual level, would be powerful auxiliary information in selecting samples for testing. As well as providing a solid weighting structure to extrapolate results to national level, it would provide a basis for evaluating and contrasting attainment according to the type of educational intervention.

As a source of indicators of outcomes in adult life, the inclusion of the PPS Number at primary stage would allow an individual's educational record to be related to employment and other outcomes in later life - either through administrative sources or household surveys in which the PPS Number is collected.

2.3.2 Second level sources - the Post Primary Pupils database

Unlike the Primary Schools data, Post Primary data are held at individual student level. Every October, each of the approximately 750 schools submits a return directly to the Department. Most of the returns are made on diskette using one of two software packages (both of which are supplied by external companies). A large amount of information is held relating to each individual student. The main variables held are student name, address, date of birth, sex, nationality (if not Irish), entry date, course or programme code, programme year and repeat year indicators. Importantly, the PPS Number is included (at the moment, PPS Numbers are not collected by the school for new entrants, but are matched with DSFA using address strings and the PPS Number is then issued to the school through the Department). About 95% of students currently have a valid PPS Number attached. Importantly, for those students who left school without completing the senior cycle a "Left Early Indicator" is attached to the student's record. As well as the date left (which is for the previous academic year) a variable indicating the destination (whether pursuing further education in another school or third level institution, gone into employment, apprenticeship, etc.) of the student is included. This is particularly relevant to Section 2.3.6, which deals with destination of exits surveys.

The system also caters for ESF funded initiatives (VTOS and VPT2). The system also acts as an application point for State examinations. Subject markers are attached to each student, with indicators of subject level for those due to sit examinations (Junior Certificate, JCSP, Leaving Certificate, and LCVP). Upon receipt of the file, examination numbers are issued - while PPS Number is not included in examination databases, a one-to-one link between PPS Number and examination number is available (provided that the year of sitting the examination is known).

The system also includes information at the level of the school: unique school number; school year; official school name and address and contact details; and Health Board Area and Community Care Area (for EHB and SHB only) - most of the school level details would remain constant from year to year.

As data entry is mainly carried out locally by the school using customised packages, there are logistical difficulties for the Department in ensuring timely and accurate returns. A very detailed manual is supplied to schools each year that highlights any changes to the system from the previous year. However, delays and quality problems arise if "the person who knows how to use the package" within the school is not available. Also, as the system is supplied by a private company, implementing year on year changes can be problematic (and expensive) for the Department.

The introduction of a Primary Pupil Database would have implications for operation of the Post Primary Database. For instance, the availability of the PPS Number on the primary database will mean that information can be carried through directly from it to the post

primary database[11]. Issues relating to educational disadvantage, disability and special or social needs (referred to in the previous section) are of course also relevant at post primary level.

2.3.3 Surveys of achievement at First and Second level

The ERC conducts a number of surveys of educational achievement at various levels of the primary and second level cycle on behalf of the Department. Among these are:

- *National Assessment of English Reading*, which is an investigation of standards of reading among pupils in 5th class. Almost 4,000 pupils in 150 schools are assessed. It was carried out in 2003 and will be repeated in 2006 (in order to measure achievement of target set out in NAPS);

- *National Survey of Mathematics Achievement*, which was conducted in 1999 among pupils in 4th class. Almost 5,000 pupils in 120 schools were assessed by means of written test. It is intended that the results of this survey will provide benchmarks to assess performance when the survey is repeated, probably in 2005;

- *Suirbhé Náisiúnta Inniúlachta sa Ghaeilge I Rang a Sé,* which was conducted in 2000 and was a study of the achievement in the Irish language of pupils in 6th class. About 4,000 pupils in about 220 schools participated; and

- *Survey of Reading Literacy in Designated Disadvantaged Schools*, a survey of about 1,200 pupils in 70 schools designated as disadvantaged (see discussion of designation criteria in Section 2.5). It was carried out in 2003 and will be repeated in 2006 (in order to measure achievement of target set out in NAPS).

The above were all conducted by means of testing with supplementary questionnaires for the school and parent. All are designed to establish the factors that influence achievement including, of course socio-economic and other factors at the level of the individual pupil. All are relied on heavily to produce many of the indicators listed as data needs by the Department.

Program for International Student Assessment (PISA)

The Program for International Student Assessment was carried out by the ERC on behalf of the Department. It was part of an international exercise, 27 other OECD countries and four other countries participated. In Ireland, almost 4,000 students in 139 schools were administered tests to assess achievement in reading literacy, mathematics literacy and scientific literacy. PISA is run on a three years cycle. In 2000 the major focus was on reading literacy, while in 2003 and 2006 the major focus will be on mathematical literacy and scientific literacy, respectively. As well as the tests, questionnaires are also administered to students with questions concerning family structure, socio-economic status of parents, and highest level of education of parents. A questionnaire is also completed by schools which collects information on the school environment.

The results of the 2000 PISA study for Ireland highlighted certain factors that affect performance. For instance, socio-economic status and highest level of educational attainment of parents were found to be significant predictors of performance. Being from a

[11] There are arrangements between primary schools and secondary schools regarding the forwarding of information at the individual pupil level. Links between the two pupil databases could formalise this transfer of information so that the flow of data would be between schools and the department.

large family or part of a lone parent family were found to have a negative influence on performance.

While aggregate findings like these are extremely valuable from a policy perspective, they bring even more into focus the need for a Primary Pupil Database, which would include indicators of the factors influencing performance. Such a database could be used as a sampling frame for testing, while simultaneously controlling for educational intervention so that measures of effectiveness of the intervention could be more precisely developed. Of course, the availability of individual databases at primary and secondary level would facilitate longitudinal testing which would be a powerful tool from the perspective of educational research.

Individual level databases (with the relevant indicators) could be used in conjunction with more widespread testing to measure the effectiveness of schools[12]. Results of testing (whether once off or as part of a longitudinal programme) could be evaluated in the context of individual pupil characteristics as captured on the pupil database.

2.3.4 State examination results

The Post Primary Pupil Database is the initial source for the identifier information held on the examination database. The pupil database contains indicators of examination entries, including information on subjects and levels to be taken. Details of pupil name, address, gender, as well as the school identifier are fed from the pupil database to the examination database which is held by the newly established Examinations Commission (formerly the Examination Section of the Department). Recently, the PPS Number has also been supplied although this is not yet used systematically.

In order to confirm medical card status, the Examination Commission issues a form to all students. This information is required to assess eligibility for (waiving of) examination fees. This information on medical card eligibility is currently the only source of socio-economic information about examination entrants. This is mentioned because it is significant that some direct collection of information at individual level is already carried out around the time of state examinations. It is relevant to Section 2.5, where the idea is proposed of collecting more complete socio-economic information at this stage of the education cycle.

Essentially, the above identifier details (with the notable exception of PPS Number) are passed to the Central Applications Office (CAO) along with examination outcomes so that applications for third level places can be processed.

Links at individual level between the examination database and the Post Primary Pupil Database are possible but tortuous. This process would be greatly simplified if the PPS Number were forwarded directly.

2.3.5 Third level administrative sources and National Survey of Access to Higher Education

As the HEA is due to assume administrative responsibility for all third level institutes from 2004, it is more appropriate to describe its planned data holdings rather than address the data holdings currently held by the Department for institutes of technology and some other third level institutes. The HEA is currently building a data collection system that will

[12] School effectiveness cannot be assessed by measuring achievements in literacy or numeracy alone. It should also be pointed out that we haven't been told by the Department that this type of measure is a current need. We mention it because it may become a departmental need in the future, especially if the statistical infrastructure exists to support it.

eventually be used by all third level institutions. The impetus for this system comes in part from a recommendation of the Higher Education Statistical Review Group. This group, chaired by the HEA, also included representatives from the Department, Forfás, NCEA, CAO, Universities and other third level institutions and ESRI.

Each institution will provide a file to the HEA (in standard format) which will include details for all students in the current and previous academic year. This forms the basis of the annual Enrolment and Graduand Report (which was a recommendation of the expert group).

Data concerning the course and programme pursued are contained in this source. Field of study is coded to ISCED three digit level. Also, characteristics for each student are recorded. These characteristics include fields indicating whether the student benefits under the free fees initiative and whether the student is in receipt of assistance towards the registration fee. No information on entitlement to maintenance grant is included. The other variables can be grouped into those relating to academic history of the student and the student's socio-economic background.

For previous academic history the following variables are included:

- From CAO, the school at which the applicant first took the Leaving Certificate examination;
- From CAO, the points score achieved at Leaving Certificate;
- From CAO, the year at which the Leaving Certificate was taken for the last time;
- From CAO, the subjects and level taken and grades achieved at last Leaving Certificate sitting;
- The last awarding institution attended and the year of leaving this institution; and
- Highest qualification attained prior to course entry.

From the above, it is apparent that certain CAO fields are passed directly to the institutions who then individually feed some of this information to the HEA. There is an obvious argument for streamlining this process - if examination results and CAO files contained the PPS Number, it would be possible for the HEA (or Department, or both) to have direct access to this information.

For socio-economic background, the following variables are to be collected:

- Principal Economic Status of both father and mother;
- Occupation of father and mother (previous occupation where relevant);
- Employment status of father and mother (i.e. whether employee, self employed etc); and
- Farm size (where relevant).

The above variables are to be collected by the institutions at registration. They are traditional Census of Population (COP) variables. This is a very deliberate approach as the capture of this socio-economic information is prompted by equality and access considerations which have been addressed up to now by the National Survey of Access to Higher Education.

National Survey of Access to Higher Education

This survey has been conducted at intervals of 6 years since 1980. The latest survey was in 1998. The target population is all first time entrants to third level educational institutions. The primary purpose of this series of surveys is to monitor participation rates by socio-demographic background[13]. There have been differences in the operation of the survey over time - for instance, in 1992 survey forms were issued with CAO forms, whereas in 1998 they were sent separately by CAO. There were almost 33,000 responses to the 1998 survey, which represented a decrease in the response rate from 1992. The change in the way the survey was administered probably accounts for the drop in response rates between the two survey occurrences.

The drop in response rates posed an additional problem for the researchers. Various analyses were carried out to check the representativeness of the 1998 survey respondents. These analyses, and indeed the whole survey process, lean heavily on the availability of "whole population" COP information. This is because the entrants to higher education constitute one side of the equation. Obviously, the other side is made up of those who don't enter third level. Essentially the survey uses COP classifications to place third level entrants in the context of the socio-economic profile of the population as a whole. Changes in COP occupation coding have caused problems for the researchers in deriving socio-economic variables, as the time series element of the surveys is an important consideration. At times, a dual-coding operation has been carried out by the researchers in order to provide a link to facilitate comparability over time.

The survey combines the direct responses from students with CAO information and some information supplied by each third level institution. The third level institutions supplied information on the course of study, county of origin and source of financial aid. The CAO information supplied was in respect of school attended. The identifier used to link the three sources was the CAO number (which was included on the form sent to students),

The data that the HEA intends to collect from each institution (as outlined in the previous section) are very similar to the data collected in this survey. Presumably this is deliberate, so that analysis similar to that conducted at six-yearly intervals could become available annually from administrative sources.

However, a lot of effort is expended in relating the population of third level entrants to the general population. This effort would not be necessary if the socio-economic background of all students at a particular stage of second level education were collected and not only information for those who enter third level institutions. As long as a unique individual identifier is available (the PPS Number) at the two stages (second level stage and entry to third level) then the output of this survey would be directly derivable from administrative sources (and annually, rather than every six years). An appropriate stage of second level education at which this information might be collected would be at the end of compulsory education - i.e. around the time of the Junior Certificate examinations. Note that PPS Number is already available on the Post Primary Pupil Database.

A further advantage of collecting socio-economic data at Junior Certificate stage is that it could feed directly into the survey of school leavers which is considered in the next section.

[13] The survey achieves more than this - for instance, participation rates are broken down by field of study, type of school attended, gender, etc.

2.3.6 Destination of exits surveys - the First Destination Report and the Survey of School Leavers

First Destination Report

The first destination report is produced by the HEA on the basis of data collected by each individual institution - usually in respect of a period about 18 months after graduation. The data collection is carried out by postal survey. The survey has been conducted annually since 1982. For the most recent survey for which results are available (2000), there were 23,000 respondents from a population of 37,500, giving a response rate of 62%.

The survey classifies respondents as to whether they are in employment, seeking employment or not available for employment. For those in employment, sector and earnings information is collected.

Non-response is clearly a serious issue for this exercise. Also, the fact that the data collection and (to some extent) survey procedures are under the control of individual institutes may be an issue.

Again, if the PPS Number were to be available from the HEA individual student database, this would open up the possibility of collecting much of the information from administrative sources. For instance, Revenue files would contain PPS Number, sector of employment, date of commencement of employment and earnings information. The DSFA sources would also be relevant. Not all of the relevant information would be available from these administrative sources so some direct data collection might still be necessary. However, this could be organised centrally and the relevant information fed back to the institutions.

Survey of School Leavers

The Survey of School Leavers is carried out by ESRI on behalf of the Department. The survey was conducted for the first time in 1980 and until relatively recently was conducted annually. The latest survey was conducted in 2003 in respect of those leaving secondary school in the academic year 2000/2001. The sample size is approximately 2,500 and the response rate is approximately 73%. Data are collected by means of personal interview (although some response is proxy - the selected individual may be unavailable due to having left the country or for some other reason). While 73% would be considered high for a survey of this nature, a lack of information for 27% of the target population is clearly an issue for a survey concerned primarily with destination of school leavers.

For the 2003 survey, the sample was selected centrally by the Department. Individuals were contacted by the Department initially, which allowed selected individuals to opt out of the survey. The population of early school leavers, those who left school before completion of the Leaving Certificate cycle, was over-sampled in order to achieve a representative response for this sub-group (because the number of early school leavers is much lower than the number who complete second level).

The survey questionnaire is long and detailed. One reason for this is that it is assumed that nothing is known about the respondent until the interview occurs. So, for instance, information about the last exam taken (including subjects taken and results), the stage of education on leaving and whether the respondent later attended another school to undertake a post-Leaving Certificate or VPT course are asked. Questions relating to third level education are also asked: whether the respondent is enrolled in a full-time or part-time course; the college or institution at which enrolled; the programme of study; whether degree, diploma, certificate or other qualification is being pursued; and whether maintenance or other

forms of grant aid is received. Other questions aimed at those working full-time – relating to the nature of the work, pay, sector, etc. are included. Detailed questions about unemployment periods are also included.

Even though the questionnaire is long and detailed, there are large segments of it that are not relevant to many respondents. This, of course, is a consequence of the approach taken - to assume nothing is known about the respondent until the interview occurs.

It is clear that reliable information on the above topics is already, or could easily be made, available. Given that the Post Primary Pupil Database contains the PPS Number, information about subsequent enrolment in other schools should already be available. Information concerning registered unemployment and all other short-term allowance or benefit schemes is already contained in DSFA sources - the CRS and ISTS. Information on employment (dates of commencement, cessation, earnings, sector of activity) is contained in Revenue files. Indeed, most of this Revenue information is already fed directly to the DSFA CRS system, so that by referencing one source only, a large amount of information relevant to some of the topics of the survey would be available.

Of course, it would also be possible to have available the subjects studied and the level and results of exams taken. Also, the capture of the PPS Number at entry to third level (by, for instance, feeding it from the Pupil Database to the examination database through the CAO system) would mean that precise information would be available to augment (or replace) this aspect of the survey of school leavers.

A section of the questionnaire is devoted to questions about Principal Economic Status (PES) and occupation of parents. These questions are almost identical to those included in the National Survey of Access to Higher Education and in the proposed student database to be introduced by the HEA. We have already indicated that there is a strong case that this information be collected at the end of compulsory school attendance (at junior certificate level) rather than be collected only for those who enter third level education. If this was the case, it is likely that those questions would be unnecessary in the survey of school leavers (as the information would already be known).

There may be an issue with regard to the role that destination of exit surveys play in the statistical infrastructure of the Department. When the Survey of School Leavers was initiated in 1980, its primary function was related to the issue of youth unemployment. Sources of unemployment data have greatly improved since then (because of the development of the QNHS and DSFA sources). It is clear that the focus of the survey has broadened but we are unsure whether this was because the survey evolved in its own right or as a response to clearly articulated needs. In any event (if the survey is to continue into the future), there is a clear case for using all of the available information for purposes of sample stratification and hence survey weighting. Using that information would also allow for more targeted questions to the various sub-groups. It could be the case that questionnaires could be customised for some clearly identifiable sub-groups.

2.3.7 General population surveys

Quarterly National Household Survey

The Quarterly National Household Survey (QNHS), which contains questions on educational attainment, is an important source for the Department in monitoring the educational profile of the population. This information can be analysed in conjunction with the large amount of information that exists in the QNHS concerning household characteristics and the demographic and socio-economic characteristics of the population. We are aware of a

current issue regarding the categorisation of educational attainment for departmental purposes. This is the subject of discussions between the Department and the CSO.

The inclusion of PPS Number in the QNHS may be desirable in the future from a departmental perspective. However, this would involve a long-term perspective, as the educational sources that capture PPS Number would have to exist before this would be useful.

Household Budget Survey

The Household Budget Survey (HBS) has been used by the Department as its only source of household expenditure on private education. As the HBS occurs only every 5 years, it may be more appropriate to consider collecting this information from the QNHS. Apart from the much larger sample size, there may be advantages for the Department in having a "social module" on the subject of private education included at regular intervals in the QNHS.

EU Statistics on Income and Living Conditions

The EU Statistics on Income and Living Conditions (EU-SILC) which is the successor to the Living In Ireland Survey, also includes educational variables and from this year will capture PPS Numbers directly from households.

International Adult Literacy Survey

The International Adult Literacy Survey (IALS) was conducted by the ERC on behalf of the Department in 1994. This was part of a co-ordinated international exercise. A sample of over 2,000 individuals was included in the survey. Comprehensive literacy testing was carried out alongside the collection of demographic and socio-economic variables. The survey is expensive and is to be repeated for many participating countries. As of now, it is unclear whether the Department intends to participate in the next stage of the programme. This is an important issue, as information from this source would be used to monitor NAPS targets, as well as being a direct input to the United Nations Development Programme Human Poverty Index 2 (UNDP HPI2).

2.4 Main conclusions and recommendations

(i) The introduction of a Primary Pupil Database is crucial and urgent. The capturing of the PPS Number is essential for coherence with other sources and to enable effective targeting of educational disadvantage. With this in mind, we give an outline (in Section 2.5) of a possible method that could be used to introduce a basic but functional database relatively quickly.

(ii) The systematic use of the PPS Number at all stages of education is essential for the administrative and statistical needs of the Department. The PPS Number should be captured at primary level and fed through the system to second level, examinations databases and third level. Schemes for other state-aided education and training (part-time education, further education, apprenticeships, etc.) should also capture the PPS Number and those sources should be available to the Department, whether or not the Department is involved in their funding. The legal situation of data collection from private schools should also be addressed so that the Department has access to information relating to all children and students.

(iii) The Department should liaise closely with DSFA to ascertain precisely the information that is available now from its CRS and ISTS systems and its potential usefulness for administering and monitoring the outcomes of the education system.

(iv) The available individual level databases should be exploited to provide the sampling frame for efficient sample selection for testing purposes. This includes testing for the purpose of longitudinal research.

(v) The Department should consider the case for direct data collection from students and households at various stages of the educational cycle. We believe there is a strong case for the collection of demographic and socio-economic information at entry to primary education. There is also a very strong case for collection of this information at Junior Certificate examination stage. If this were done, the National Survey of Access to Higher Education would be redundant.

(vi) The role that destination of exit surveys play in the statistical infrastructure of the Department should be re-appraised.

(vii) The Department should be the ultimate owner of the micro data for all surveys that it funds.

(viii) Information on the data sources, and associated metadata, owned by the Department should be clearly publicised both internally and externally.

(ix) The development of coding frames for disability and special or social needs as well as ethnicity should be pursued with CSO, DSFA and the Equality Authority.

(x) The Department should begin an iterative process to identify more precisely its information needs and the sources it intends to develop to meet those needs. This should be an iterative process as needs are sometimes expressed in the context of available sources and therefore Department needs will develop as better data become available.

2.5 Additional proposals

Designation of disadvantaged school status

As the issue of educational disadvantage is highly important for schools at both first and second level, we give a brief outline of the current methods used by the Department in identifying disadvantage.

There are currently three types of scheme:

- Schemes such as "Giving Children an Even Break" (GCEB) which are aimed at schools containing even a small number of disadvantaged children. Under this scheme, the school identifies qualifying children and additional resources are allocated to the school. The amount depends on the number of qualifying children.

- Schemes such as Early Start and the School Completion Programme which are aimed at schools that have a proportion of disadvantaged students which is above a certain threshold.

- Schemes based on the area of the school. These schemes are usually part of cross- departmental area-based initiatives.

The identification of educational disadvantage has been achieved in the past by means of questionnaires to school principals. For many schemes, this involved the principals estimating the number of children whose families are entitled to medical cards and reporting these aggregate figures to the Department. Within the Department, concern has been expressed about the appropriateness of this measure as well as the consistency of reporting from school to school.

There is a clear case for using a more consistent and appropriate measure of socio-economic disadvantage. The availability of a Primary Pupil Database with the PPS Number identifier is central to the development of such a measure. There are two reasons for this:

- It would allow direct linking at individual and family level to the DSFA Central Records System and other systems (e.g. ISTS) which feed directly into the CRS.

- It would provide a unique pupil identifier, with home address, which could be used for issuing questionnaires to households.

The latter point would be particularly relevant if sufficient information could not be accessed via DSFA sources, or if it were deemed necessary to collect other background information for families (e.g. highest educational attainment of parents).

Because of the importance of the Primary Pupil Database, we give a broad outline here of a possible method to collect PPS Numbers of pupils which might be established relatively quickly.

Introducing a Primary Pupils Database

A project team from within the Department is currently undertaking a requirements analysis for a Primary Pupil Database, the introduction of which is, as has been stated earlier, a departmental priority. The results of the analysis will be used in the development and implementation of the database as part of the Educational Services Interactive (ESI) Project, which is due to begin in late 2003.

The objectives are:

- To gather valuable, reliable and timely information about primary school children and primary schools;
- To minimise the number of requests by the Department to schools for information; and
- To provide schools with a web-based application to input the required information.

However, full database implementation may be some time away and, accordingly, we put forward here a possible interim solution.

There are almost 3,300 primary schools in the state. Each of these schools will be supplying the individual level information that will feed into the Primary Pupil Database. It would seem, therefore, that a system that would allow schools to easily compile and transfer this information to the Department is desirable. The easy compilation of the information would involve familiarity with the software package that is used. Consideration should be given to using spreadsheets as the initial source of data capture at the school level. The system could work as follows:

1. Each school would list (by class or standard) the name of all pupils in the school in one column of the spreadsheet. Each row would then become a record for an individual pupil. The record would include PPS Number, date of birth, address and any other relevant pupil characteristics. Other relevant characteristics might, for instance, include a field for special needs or an indicator of socio-economic background. Those fields could be filled using a "drop-down" menu coding system attached to each field.

2. The spreadsheet could then be transferred to the Department electronically either by encoded e-mail or via a departmental "secure deposit box" using the internet.

3. Upon receipt of the spreadsheet file, the Department would open the file and import the fields directly into a common statistical analysis package (there are many available). A customised edit program could then be run on the data from the school to check for obvious errors. The school could be alerted immediately to any errors or problems with the file so that it could be re-submitted quickly.

4. The data from each school could then be aggregated (within the same statistical package) to provide one file for all schools and all pupils. It could then be fed into a longitudinal database of whatever design the Department chooses.

Clearly there will be longitudinal considerations in the way the data is held in a central database (addresses can change, names can change, etc.). However, this would be the

case no matter how the data is to be collected from the source schools. Provision can be made for year to year changes within the source spreadsheet.

Some advantages of such a system are given below:

- All schools have PC's and most, presumably, already have common spreadsheet packages. It is likely that there would be a broad level of understanding of spreadsheet packages already available in most schools. The training requirements (which might be very substantial using alternative approaches) could therefore be minimised. For the Post Primary Pupils Database (returns for which are supplied by about 750 schools), timeliness of the returns can be dependent on the availability of the (often one) person who has used the software previously. With a spreadsheet approach, anyone with a basic knowledge of common spreadsheet applications could complete the annual return;

- Class lists could be retained and used for other school purposes. Also, rollover from year to year would be straightforward. In fact, one could envisage a scenario whereby the task of maintaining the lists could be delegated (a list could be maintained for each class) and the final return could be an amalgamation of these separate lists. In other words, it wouldn't be necessary for schools to wait on the issue of "this year's" version of a custom written software package;

- A high level of pre-formatting of the spreadsheets could be achieved. For instance, the Department could make available annually (on its web site) a blank preformatted spreadsheet that could be downloaded and completed by each school;

- A significant degree of flexibility could be achieved with control of the process residing entirely within the Department. For instance, fields could be added/deleted from year to year reflecting needs as they arise. There would be no need to involve outside consultants in implementing changes. We do appreciate, however, that the key identifier fields should be kept as constant as possible from year to year - as an aid to schools, if for no other reason; and

- From time to time, information at the level of the school is required. This could be incorporated as part of the annual return simply by using another worksheet within the spreadsheet. Within this, drop down coding frames for categorical questions could be used.

A possible disadvantage (although it is debatable whether it is indeed a disadvantage) would be that the skills and resources would need to be available within the Department to operate such a data collection system. However, the skills required are mainly those that would normally fall into the category of end-user computing and involve some spreadsheet knowledge as well as some knowledge of a statistical analysis package.

Direct collection of socio-economic and other indicators at an individual level

Once an individual primary database with PPS Number is available (using whatever data collection system is deemed appropriate), the possibility of centrally collecting socio-economic and other data at the individual pupil level is greatly facilitated. This collection could be carried out at entry to primary level or at any other stage (see recommendation regarding the appropriateness of data collection at entry to the Junior Certificate examination). One data collection model would be to print questionnaires which are

customised (name of pupil and PPS Number could be printed on each questionnaire). The PPS Number would then become the key so that each completed questionnaire could be scanned upon return and the information captured with a minimum of manual intervention. Questionnaires could be issued through schools or directly to home addresses. Depending on the type of information collected, the questionnaires could be returned either directly to the Department or through the school.

Chapter 3

Department of the Environment, Heritage and Local Government

3.1 Introduction

The functions of the Department of the Environment, Heritage and Local Government (DoEHLG) are wide-ranging and include Housing, Roads, Environment, Heritage, Local Government, Planning, Water Services, etc. Each of these areas contains sources of information derived principally from administrative records. There are also a number of public bodies operating under the aegis of the Department, such as An Bord Pleanála, the Environmental Protection Agency, the National Safety Council, An Chomhairle Leabharlanna, Comhar, the Dublin Docklands Development Authority, the Dublin Transportation Office, the Local Government Computer Services Board (LGCSB), the Fire Services Council, the Housing Finance Agency, etc.

The Steering Group on Social and Equality Statistics (SGSES) identified 24 data holdings held within the DoEHLG or in Local Authorities (LAs) or other bodies operating under its aegis. Two-thirds of these data sources relate to the various housing schemes administered by the Local Authorities. The remaining sources related to the House Price system, the National Survey of Housing Quality (conducted for the department by the ESRI), the National Vehicle & Driver File, the Register of Electors, Fire Statistics, and the EPA's annual "Environment in Focus" report.

The DoEHLG SPAR team considered these and other sources in the light of the data needs identified in the SGSES Report. Given the tight time constraint for the SPAR Project, the team decided to concentrate on the following sources:

- The new Integrated Housing System (which is subsuming the bulk of the data sources identified in the SGSES Report);
- The new House Price system being put in place;
- The planned Private Residential Tenancies Register (a new source not listed in SGSES);
- The National Vehicle & Driver File; and
- The Register of Electors.

These prioritised DoEHLG data sources are discussed in the following sections.

3.2 Integrated housing system

While the DoEHLG is responsible for policy and legislation in the area of housing, the services are virtually all delivered through the local authorities. They include:

- Provision, maintenance, and management of local authority housing;
- Provision of traveller accommodation;
- Delivery of accommodation services for homeless people;
- Support for the provision of social housing projects by voluntary housing bodies;
- Improvement works to privately owned houses in lieu of local authority housing;
- Shared ownership;

- Affordable housing;
- Sale of local authority houses to tenants;
- Loans for house purchase and improvement;
- Low cost housing sites;
- Mortgage allowance scheme;
- Disabled persons grants; and
- Essential repairs grants.

These services involve a wide range of schemes, and each authority has its own computer system to manage these schemes. However, under the Integrated Housing Project, which is underway at the moment, the aim is to computerise all the housing schemes using a common system (and, probably, a single database). The project is fairly advanced and a system is being commissioned by the LGCSB. A pilot implementation will be carried out in one particular local authority later this year, though it is expected that it will be 2005 before there is a full roll-out of the product. Application forms, which currently differ across the LAs, will be standardised. Data will be recorded by each LA, and the LA will only have access to its own data. The DoEHLG will have sufficient access to the database to generate the reports they need.

The system will hold information on people, property, rents, repairs and allocations. Information on people will include PPS Number, sex, date of birth, marital status, and relationship to head of household. The income of individuals and households will be captured[14]. Some information will be collected on occupation (nine categories) and employment sector. Nationality (not ethnicity or place of birth) will be collected. It will also be possible to see who's on the housing list, the rehousing/transfer list, the Traveller list, and the homeless list.

The properties to be included in the system will be those constructed, owned, leased or rented by the LAs, whether occupied or not, including Traveller halting sites and halting bays. Currently, there are about 100,000 occupied LA dwellings in the State. The system will capture many property attributes, such as acquisition date, heating/insulation/plumbing, parking, number of bedrooms, etc. The system will not include details on private properties, or B&Bs (apart from addresses), or properties operated by voluntary associations, even if they are used to house LA applicants. However, details will be included on the LA clients housed in such properties. It has not yet been decided if the Geographic Information Systems (GIS) co-ordinates of properties will be held in the system. Currently, it is not planned to use the An Post/Ordnance Survey Ireland GeoDirectory in the project.

There are essentially 88 LAs, though it is unclear yet how many authorities will be distinguished in the new system. Some Local Authorities split the housing function into a few housing authority areas. However, it will be possible to aggregate the data to the level of County.

Of the 24 SGSES sources relating to DoEHLG, the following 16 are being subsumed into the new integrated system:

[14] Applicants will have to supply their P60 where relevant.

- Affordable Housing Scheme;
- Affordable Housing Scheme Loans (new houses only);
- Annual House Letting Survey;
- Assessment of Need for Social Rented Housing Accommodation;
- Homeless (as part of triennial Housing Needs Assessment);
- House Purchase Loans;
- Housing Loan/early redemption/special payments;
- Mortgage Allowance Scheme;
- Rental of LA dwellings;
- Sale of Local Authority Housing;
- Sale of Site scheme;
- Shared Ownership Loans;
- Shared Ownership Scheme;
- Tenant Purchase Loans;
- Tenant Purchase Scheme - LA returns to Housing Policy and Supply, DoEHLG; and
- Traveller Accommodation (Annual Count).

As this is a new system being put in place, it is not possibly to assess the quality at this stage.

3.3 House price system

The situation here is also in the process of change. Currently, the DoEHLG collects quarterly returns from mortgage lending institutions, local authorities and the Central Bank, which it uses to compile the house price series published in the quarterly Housing Statistics Bulletin. The returns provide information on the aggregate value and number of loan approvals for each lending institution. Data are not separately provided for each mortgage transaction. Average prices are derived from the loan approvals data for all dwellings, and for houses and apartments separately, for the country as a whole and for the five main urban areas (Cork, Dublin, Limerick, Waterford, Galway). The derived house price data do not take account of changes in the mix of dwelling types. The Department also publishes in the Bulletin an annual profile of borrowers based on voluntary returns from the individuals taking out loans. The forms, which are returned by the lenders, provide information on borrowers in relation to house prices, previous tenure, marital status, income and occupation.

Under a new system, now being developed, the mortgage institutions will provide the department each month, by the middle of the following month, with an electronic file of

individual (though anonymised) transactions, giving details on the borrower, the property, and the loan. A record will be forwarded both at loan approval stage and at first drawdown (mortgage commencement) stage. The new system, with its improved frequency, detail and timeliness, will provide useful new statistical information on borrowers and will allow the compilation of a new mix-adjusted house price index. All relevant lenders will provide the information under the Housing (Miscellaneous Provisions) Act, 2002. An added incentive for the lenders is that they will have access to an anonymised version of the micro data database for analysis, with some details (including the identification of the lending institution) omitted. As the new file specification required from the lenders calls for changes to their computer systems, including mortgage application forms and data capture processes, the collection of various variables has been prioritised by the department in order to ease implementation. While it is expected that the system will be operational from the start of 2004, not all variables will be collected from that date, and the quality of data will have to be tested prior to any publications or analysis.

For the main and second borrower, the system will collect age, sex, marital status, employment status, employment sector, income information, and current tenure (type and location). For the property, the information will include the purchase price, new/second-hand, location (county, postal code), year of construction, dwelling type, construction type, area, number of rooms, details of additional costs, etc. Details will also be collected on the loan.

Geographic coding for the location of the current tenure and the location of the property will be at the level of administrative county, with postal code areas being distinguished in Dublin. It is also hoped that there will be sufficient information to distinguish between urban and rural locations.

While the lenders themselves have the PPS Numbers of the borrowers, they will not include them on the transaction file forwarded to the department. The Act states the information the Minister seeks from the lending institutions “shall not include information that directly identifies or may identify a person and, in particular, it shall not include the names, dates of birth or addresses of persons”.

The system will not collect information on non-mortgage based transactions, which are estimated to cover about 29% of all transactions. However, relevant information pertaining to the transfer of property is collected by the Capital Taxes Branch of the Revenue Commissioners for stamp duty purposes. This may be a useful source of information on properties that are purchased without a mortgage (see section 7.3).

It is likely that the data will be complete (all mortgage-based transactions will be covered) and that data quality will be high.

3.4 Private Residential Tenancies Register

The report of the Commission on the Private Rented Residential Sector, published in 2000, proposed the establishment of a Private Residential Tenancies Board (PRTB), whose main functions would be to provide a mediation service and to maintain a register of all tenancies. (In theory, landlords are obliged under current legislation to register tenancies with the relevant local authority - in practice no more than 20% of properties have been registered.) The Government set up the PRTB on an ad hoc non-statutory basis in October 2001, and published the Residential Tenancies Bill, 2003 in June. It is expected that the Bill will be enacted before the end of 2003.

From a SPAR perspective, interest is concentrated on the proposal (likely to become reality at the start of 2004) to set up and maintain a single central register of every private

residential tenancy in Ireland. It is not known how many such tenancies are in existence but it is estimated to be about 150,000 nationally. Each tenancy registered will be given a unique registration code. That code number will apply to the particular tenancy (i.e. letting) and will remain valid while the tenancy remains in existence. However, each tenancy will automatically terminate after four years, and a new tenancy may be commenced. Associated with each tenancy will be details of the landlord, as well as details of each tenant[15].

The system is being computerised at the moment, using a system called RACTS (Registration and Case Tracking System). A paper registration form will apply, followed by manual checking and manual data entry. Electronic entry systems (e.g. web registration) will be considered in future. The registration data required in respect of each new or updated registration will include:

- Information on Tenancy: Date and terms of tenancy (commencement date, period of tenancy, rent amount, frequency of payment); Registration date;
- Information on Property: Address; Name of authority where dwelling is located; Description (bedsit, apartment, terraced, semi-detached house, etc); Number of bedrooms; Number of bed spaces; Floor area;
- Information on Landlord: Name and address; PPS Number; CRO Number, if the landlord is a company; and
- Information on Tenant(s): Number of occupants; Name; PPS Number.

The geographic classification will be based on Local Authority regions, with postal codes in Dublin.

The value of the new register as a data source will largely depend on its completeness. There will be a legal obligation on landlords to comply, and the PRTB states that it will be in the interest of a landlord to register and anticipates that the compliance rate will be high. The Board's policy will be to prosecute landlords who do not register.

3.5 National Vehicle & Driver File

The National Vehicle and Driver File (NVDF) consists of a very complex database, with details of all registered vehicles and all licensed drivers in the country. It is supported and maintained in the office of the Vehicle Registration Unit (VRU) of the DoEHLG in Shannon. All vehicle and driver licence transaction processing that takes place in the Motor Tax network, and change of ownership processing in VRU, is facilitated through the NVDF, and the database is updated in real-time. The database is also used by the Garda Síochána, the Revenue Commissioners (they conduct the registration process, which collects the VRT, and details are transferred from Revenue's motor registration database to the NVDF), the Department of Transport, the Department of Social and Family Affairs (e.g., for anti-fraud purposes), the NCT, and the CSO.

Conceptually, the NVDF can be regarded as consisting of two separate databases - one of vehicles and one of drivers.

[15] Registered tenants, being those with whom the landlord agrees the tenancy, do not necessarily correspond to occupants. If, say, a family unit rents a property, then it is likely that only a single adult from that unit would be classified as the tenant.

Vehicle records are complete on the live database back to 1993, and are retained for vehicles that are no longer under current licence. The number of vehicles under licence is about 1.8 million. The primary database key is the registration number of the vehicle, which is unique and never changes. If a vehicle is scrapped, it is marked as such only if the VRU receives formal notification of the fact. Under environmental legislation, a new system, based on the use of registered scrappers, will be put in place: the scrappers will be obliged to forward to the VRU details of all scrapped vehicles, and the vehicle records will be updated accordingly.

There are many variables on the database giving details on the vehicle: make/model, year of manufacture, engine CC, Certificate of Roadworthiness, name of insurance company, number of previous owners, NCT pass and due dates etc. Details of NCT failures are not recorded. Limited ownership details are also part of the vehicle record. These consist of first name/surname or company name, address, and phone number. No other owner details are collected. In about 30% of cases the owner is a business; this could be a registered company, a private company, or a partnership. The possibility of asking for the CRO number for registered companies will be considered, but it would not be easy to apply a similar solution for unregistered companies. The complete ownership history is retained in the database. The County designation in the file is based on the Motor Tax divisions of the State. So, for instance, Dublin is treated as a single county while Limerick is treated as two counties. One of the fields indicates whether the vehicle is exempt from paying motor tax by virtue of its use for disabled drivers or passengers. In this case the vehicle must be registered to an individual and not a business.

Driver records are complete on the live database back to 1989, and the NVDF contains about two million people with a current driver licence (including provisional). The primary key is the Driver Number, which is unique and never changes. The PPS Number is collected and captured, but only from 2001. The provision of the PPS Number is mandatory – it is not possible to get a licence unless a valid number is provided. As driver licences can last for up to ten years, it will essentially be late 2009 before all drivers have a PPS Number on the system. The first exposure of an individual to the system occurs when he or she takes the Theory Test, which must be successfully completed before a Provisional Licence can be obtained.

Apart from the reference numbers and the name and address details, the driver information includes date of birth, country of birth, sex, details of endorsements or disqualifications, and the answers given to 11 yes/no health and fitness questions; these questions relate to eyesight, hearing, need for vehicle modification, etc. Some special restrictions may be held as special codes on the physical licence, such as 01 (must wear corrective lenses), or 40 (modified steering). Details of the Certificate of Competency are held, but no details leading up to the receipt of the certificate (i.e., no details of Driving Test failures). Penalty Points are recorded on the system. Currently, about 38,000 drivers have Penalty Points. The administrative action that triggers the updating of this field in the NVDF occurs when the driver turns up with his/her driving licence to pay the fine involved.

There is no explicit connection between vehicle and driver details. While it would often be possible to identify the driver of a vehicle by doing a search on name and address, there is no link between the name and address of the owner on the vehicle record and the name and address of the driver on the driver record. Indeed, they will often be different because the owner is not the driver, for example in the case where a company is recorded as the owner of a range of cars. There are no plans currently to collect the PPS Numbers of owners on the vehicle record.

The quality of the NVDF data is considered very high. Timeliness is not an issue – the database is kept up to date, with records being added and updated daily. There are no

problems with duplicates. The quality of the address information is high for vehicle owners: a sample tested automatically by An Post showed that 98% of the addresses were valid. Moreover, of over a quarter of a million notices posted out each month, less than a thousand are returned with address unknown.

To obtain his/her first provisional licence, a person must produce a birth certificate (or, for non-EU applicants, a passport). This is used to check the date of birth on the application form. It is also used to check the person's name, and documentary evidence must be produced if the names are different.

3.6 Register of Electors

The Register of Electors is used for Presidential, Dáil, European and Local elections. Conceptually a single register, it is in fact a set of separate registers maintained by the registration authorities. Policy, co-ordination and legislation are handled by Franchise Section, DoEHLG.

Each of the 34 Registration Authorities - there's one for each administrative county - maintains its own register. While each register is separate, they are all computerised using the same LGCSB system. A consolidated version is not yet available or being planned, but may become available at some stage.

Each authority updates its register each year. This is based on a simple application form. Apart from name and address, the form asks for date of birth and citizenship (distinguishing the four categories Irish, British, Other EU, Non-EU). Date of birth, the collection of which only started last year, is not given on the published register, but it is captured in the computer system. After the local authority publishes the register in February of each year, anyone not on the register may apply for inclusion in the Supplement. At the time of an election, the persons entitled to vote, depending on their citizenship type, are those on the main register as well as those on the supplement a specified number of days before the election. The reason for the two registers is purely administrative, and in practice the register for a particular authority may be considered to be the two combined. The number on the supplement would amount to only a few thousand - a tiny fraction of the number on the register.

The register cannot be considered complete, and there is no legal requirement to register. Currently, the combined registers have a total of 2.9 million records, which exceeds the number of persons aged 18 and over in the country. Some of these records relate to deceased persons, and some are duplicates.

The register has two particular subsets. The first is the Postal Voters List, which includes whole-time members of the Defence Forces and Irish diplomats posted abroad and their spouses. The following may also apply: members of the Garda Síochána; persons living at home who are unable to go to a polling station to vote because of a physical illness or disability; and persons whose occupations are likely to prevent them from voting at their local polling station on election day, including full-time students registered at home who are living elsewhere while attending an educational institution in the State. The second subset is the Special Voters List. This consists of electors with a physical illness or disability living in hospitals, nursing homes or similar institutions who wish to vote at these locations. In order to qualify, an elector's physical illness or disability must be likely to continue for the duration of the register and prevent him/her from going to a polling station to vote. Neither of these subsets can be considered complete. For instance, a Garda may decide not to apply for the postal option, or a person in an institution might not apply for the special list. Information on occupations or disability is not held on the computer system - instead the system maintains a

flag on a record if it is part of the postal list or the special list, and this translates to a special letter code on the published register. The total number of postal voters is about 13,000 and the total number of special voters is about 3,000.

A company called Precision Marketing Information Ltd (PMI), an associate company of An Post, takes the existing registers, combines them, adds household identifiers (by merging addresses), and eliminates some of the duplication. The resulting database is sold commercially for creating national mailing campaigns, for generating samples for surveys, etc. The CSO retain the services of PMI annually for this database - it is used in the Household Travel Survey for sampling, and, in part, for grossing the results. PMI also used the database as a basis for its own lifestyle survey, which went to over a million households. The resultant detailed lifestyle database of over 250,000 records is also sold commercially. Recent statistical legislation limits what PMI and others can do with the register. The application form for the Register of Electors now contains the question “Do you wish your registration details to be available for use for a purpose other than electoral or other statutory use?”. In future, registration authorities will publish two versions of the register - the edited register and the full register. The edited register will contain persons who have indicated that their details can be used for other purposes. The full register will list everyone who is entitled to vote. When the edited version is published, the full one may only be used for an electoral or other statutory purpose.

The SPAR team was informed of the following data protection issue that arose. The DoEHLG wanted all persons reaching the age of 18 to receive, automatically, an invitation to join the Electoral Register. So they asked the Department of Social and Family Affairs to provide the relevant names and addresses to them. This was refused, citing data protection concerns.

3.7 Other data sources

This section briefly mentions some sources in the SGSES Report that the SPAR team did not investigate in detail.

Fire Statistics

DoEHLG publishes annual fire statistics, which gives an analysis of call-outs, fatalities, etc. The statistics are based on information provided by the individual Fire Authorities, notably using the Fire Fatalities Report. This form is completed by the Fire Services if they are called out and if there are deaths caused by fire. The information collected is: date/time of fire; deaths by sex and age group; type of premises/location; cause of fire; origin of fire; whether a smoke alarm is present. The Farrell Grant Sparks Report on the *Review of Fire Safety and Fire Services in Ireland*, published in March 2002, sets out recommendations for the future development of the service. The consultants examined all matters relating to the fire service, including the collection of statistical data. In this regard they identified lack of comprehensive statistical data and lack of proper research and analysis as an area that needs to be developed. One of the findings in the report was that “there are no definitive fire death statistics (there are only statistics for incidents to which the Fire Service responds), information on non-fatal fire injuries is extremely patchy, there is no reliable information on the causes of fire, there is no available assessment of the cost of damage by fire and no information on the savings accruing from Fire Service intervention.”

National Survey of Housing Quality (NSHQ):

This survey was carried out in 2000/2001 by the ESRI for the DoEHLG with a sample of 37,500 households. It is normally carried out every ten years. This survey provides detailed information on the housing stock in Ireland. The main purpose is to provide DoEHLG with an assessment of the current national housing stock and the condition thereof. It will facilitate the analysis of results for households such as large families, lone parents, elderly persons living alone, etc. The ESRI will forward the database of anonymous confidential data to the DoEHLG. The PPS Number is not collected. The survey was previously called the House Conditions Survey and was undertaken by Housing Inspectors, who assessed the properties as part of the survey. The latest survey was conducted by interview, so the answers are based on the perceptions of residents and tenants of the conditions of the properties.

This survey is more detailed than the QNHS housing module and generally includes all questions on the module. The QNHS housing module was undertaken in 1998Q3 and will be undertaken again in 2003Q3. In considering the cost of the QNHS module and the NSHQ, and looking at the purposes they serve, questions must be asked as to whether there is a more efficient way to fulfil these purposes. Is there a need for a QHNS module every five years? Should the timing of the QNHS module be altered to five years after NSHQ? Could the QNHS incorporate the NSHQ, or would there be too much detail?

Environment statistics

The Environmental Protection Agency (EPA) monitors the environmental health of the State and publishes reports on the state of the environment. A range of environmental indicators is available, each of which measures some aspect of the state of the environment. The annual publication *Environment in Focus 2002* contains 50 indicators, covering water quality, waste management, air pollution, climate change, urban environment, transport, biodiversity, eco-efficiency, etc.

3.8 Statistical potential

Integrated Housing System

Statistics derived from the systems currently in place (see the department's quarterly Housing Statistics Bulletin) are based on aggregate returns made by the 88 LAs to the department. Further breakdown or cross-classification is not possible. But the new system will be a database of micro data common to all LAs, consistently maintained according to common standards, and will therefore have considerable potential. The ability to cross-classify the microdata (using age, family structure, location, income, nationality, etc) will provide much more useful statistical information on tenants, households, and schemes. The database of tenants and those on the various housing lists could be used as a sampling frame, permitting surveys of marginalised groups.

House Price system

The new system, based on a comprehensive database of micro data, will permit new useful statistical information and cross-classifications, such as analysis by type of property, house price by location (county, or postal code in Dublin), age/income profile of borrowers, relationship between income, loan and house price, information on joint purchases, etc.

Private Residential Tenancies Register

The system should be able to provide useful statistical information on the rented sector, broken down by local authority area (and by postal code area in Dublin), such as average rent by property type, number of rented units by bed spaces or property type, number of landlords with multiple/single properties, average percentage change in rent, etc. It might also be possible to construct an index of the cost of private tenancies by accommodation type, size (bed spaces), floor area and location. The Register also has potential use as a sampling frame. If the register is complete and kept up to date, it would be of considerable benefit to the CSO in the areas of National Accounts (to improve the accuracy of the imputed rents) and Consumer Prices (for monitoring residential rents).

National Vehicle & Driver File

As it stands, this data system is already a source, directly or indirectly, of published statistics. Most of the tables in the Department's annual "Irish Bulletin of Vehicle and Driver Statistics" are sourced from the NVDF. These give statistics of vehicles under Current Licence, of vehicles licensed in the year, and of current driver licences. The NVDF is occasionally used to perform analyses on foot of an external request. The database is also used by the CSO for the monthly compilation of new vehicle statistics, and as a sampling basis for the Road Freight Inquiry. New statistical information should come from the analysis of penalty points.

Register of Electors

The register is analysed statistically and the results published. As it stands, it has little additional direct statistical potential. However, it may have additional use as a register source, despite its lack of classification information such as sex or marital status. The date of birth is available, but is only collected since 2002.

3.9 Conclusions and recommendations

(i) **Integrated Housing System:** With the system being put in place, the opportunity should be taken at the testing and implementation phases to ensure that all information is collected completely and accurately. For properties, it is recommended that the GIS co-ordinates of properties be recorded and held in the system. This could be done by using the An Post/Ordnance Survey Ireland GeoDirectory, or otherwise.

(ii) **House Price System:** The cross-cutting potential of this new source would be realisable if the lenders could include the PPS Numbers of the borrowers on the files they send to the Department. If it was decided from a policy perspective that the PPS Number was in fact required, (after consideration of issues such as data protection), it would appear to require an amendment to Housing (Miscellaneous Provisions) Act, 2002.

(iii) **Private Residential Tenancies Register:** Every effort should be made with this new system to ensure that the register is as complete as practicable, that information on each tenancy is complete and accurate, and that the register is kept up to date.

(iv) **Register of Electors:** The inclusion of the PPS Number as part of the registration process should be considered. It would confer administrative advantages: improved

verification at registration stage; handling of address changes; elimination of duplicates; elimination of records of deceased persons; overcoming the problem of people at the same address having the same name (e.g. father and son). But it would also, of course, confer statistical and policy advantages: for example, the register could be analysed to examine the enrolment rate for young adults.

Chapter 4

Department of Health and Children

4.1 Introduction

In the preparatory work for its report "Developing Irish Social and Equality Statistics to meet Policy Needs", the Steering Group on Social and Equality Statistics (SGSES) identified 24 potentially relevant individual data holdings administered either directly within the Department of Health and Children (DHC) or by bodies operating under the Department's auspices. In the context of the terms of reference of the SPAR Project, these sources, along with the information collected by the General Register Office, were to be the focus of attention. The main objective of the Project was to examine the possibilities for enhancing the statistical potential of existing data rather than how such enhancement might be achieved or who would undertake the necessary tasks. The DHC SPAR team was also conscious of the information needs for policy purposes identified by SGSES and was hopeful that its investigations would determine, at least in some cases, whether these could be addressed.

Much of the information, which the DHC holds or has access to, is derived from primary data collected or reported at local or regional level (by health service providers, regional health boards, etc). For the purposes of undertaking the examination, the team met officials in the Information Management Unit of the DHC and this contact was maintained on an ongoing basis throughout the project. Given the time and resource constraints, it was decided that the focus of the examination would centre on a selection of the data available, whether directly or indirectly, to the DHC. The selected sources covered a variety of types of information, for example: registrations of vital events (births, marriages, deaths), information required for the operation of the payments system for the Medical Card Scheme and the Drugs Payment Scheme, data on the physical and mental health of the population, statistics on the interaction of patients with hospital services. They also included national survey-based data designed to collect information on health behaviours and social and environmental conditions. These surveys, usually repeated periodically, are generally carried out by outside bodies on behalf of the department.

Apart from a specific contact made with the ERHA concerning one particular data source (see Section 4.2.4), the team did not have any contact with health boards or regional or local bodies concerned with the health service provision. However, as a great deal of the information available to the DHC is compiled outside the department, contacts were made with officials in bodies operating as executive units of DHC (e.g. the General Register Office – GRO[16]), and officials working for agencies in close formal or informal co-operation with the DHC (e.g. the National Cancer Registry of Ireland – NCRI[17], the National Disease Surveillance Centre – NDSC[18], the General Medical Services (Payments) Board – GMSPB[19], the Health Research Board – HRB[20], the Economic and Social Research Institute – ESRI, the Department of Health Promotion of the National University of Ireland – Galway). These contacts were made via meetings, telephone calls or e-mail.

To be aware of the broader policy context in health, the team familiarised itself with a number of key recent strategic documents and reports. These included the Government's Health Strategy (*Quality and Fairness: A Health System for You*), the Brennan Report (*Report of Commission on Financial Management and Control Systems in the Health*

[16] The GRO is the central civil repository for records relating to vital events (births, deaths and marriages) in the Republic of Ireland.
[17] NCRI: The Registry was set up under statute, and is wholly funded by the Department of Health and Children. The overall direction of the Registry is the responsibility of the National Cancer Registry Board.
[18] The NDSC was set up in 1998 conjointly by Ireland's seven Health Boards along with the ERHA and with the approval of the Minister for Health and Children.
[19] The GMSPB is a joint Board composed of officers from the regional health boards and the DHC.
[20] The HRB is a statutory body that promotes, funds, commissions and conducts medical, epidemiological and health service research in Ireland.

Service), the Prospectus Report (*Audit of Structures and Functions in the Health System on behalf of the Department of Health and Children*) and the recent Government proposals on restructuring the health service (*The Health Service Reform Programme)*. In addition, the team was also broadly aware of the matters currently being dealt with in preparing the National Health Information Strategy (NHIS) which, according to the DHC, will be published shortly. The published reports clearly state the need for better based, more robust and more standardised health information and they prepare the ground for the reforms and improvements which are expected to be advanced by the NHIS.

In pursuing its main objective, the team was anxious to highlight some specific ways in which the existing health information sources could be improved to better serve wider statistical requirements of the health sector including data linkages. It was conscious of the possibility that any changes that may occur in the structure and functions of the overall health service system would have important implications for collecting and compiling statistical information. However, the team is of the view that implementation of the recommendations of the reports referred to above, will have the positive outcome of adding value to the vast body of information which currently exists within the DHC and its bodies. Moreover, the fact that health issues cut across a number of government departments underscores the need for a solid basis for the integration of statistical information over a number of statistical domains.

4.2 Examination of individual data sources

At the outset of the project, the DHC SPAR team and the DHC identified 12 (out of 24) specific data sources for priority examination along with data compiled by the GRO. The number of data sources considered was largely determined by tight time constraints for the project, while the basis for each source selection was determined by both perceived data value and potential for enhancing that value. The team had the intention of examining additional sources if time permitted, but in the end it was only possible to look at one further data source. The table below lists all fourteen data sources examined along with the names of the associated data holders.

Table 4.1 Data sources and data holders

Data source	Data holder
Civil Registry of Births, Deaths and Marriages	General Register Office (GRO)
General Medical Services Database for Primary Care Reimbursement	General Medical Services (Payments) Board (GMSPB)
National Perinatal Reporting System	ESRI/DHC
ERHA Register of Congenital Anomalies	ERHA, Department of Public Health
Hospital Waiting Lists	Department of Health and Children (DHC)
Hospital In-Patient Enquiry Database	ESRI/DHC
National Cancer Registry of Ireland (NCRI)	National Cancer Registry Board (NCRB)
National Breast Screening Programme (NBSP)	National Breast Screening Board (NBSB)
National Intellectual Disability Database	Health Research Board/DHC
National Psychiatric In-Patient Reporting System (NPIPRS)	Health Research Board/DHC
Notifiable Infectious Diseases Database (NIDD)	National Disease Surveillance Centre (NDSC)
Survey of Long-Stay Units	Department of Health and Children (DHC)
Survey of Health Behaviour in School-aged Children (HBSC)	National University of Ireland – Galway, Department of Health Promotion
Survey of Lifestyles, Attitudes and Nutrition (SLÁN)	National University of Ireland – Galway, Department of Health Promotion

The descriptions that follow have not been assigned a particular ranking order. The data sources described first relate to events or transactions concerning the entire population, followed by those that relate to health concerns of subsets of the population interacting with particular branches of the health system. The last two data sources (HBSC and SLÁN) relate to information collected periodically by way of sample survey on the health behavioural and attitudinal characteristics of people.

4.2.1 Civil Registry of Births, Deaths and Marriages

This information which is held by the General Register Office (GRO) concerns the registration of vital events in Ireland in a central civil repository. The civil registration process has been paper-based since its inception about 150 years ago but efforts have been underway for some time to modernise the system through computerisation. Progress, however, has been slower than originally anticipated but the availability of an electronic search facility for births, marriages and deaths records from 1 July this year at GRO headquarters is a welcome advancement. Records currently relate to births, marriages, deaths and adoptions in the Republic of Ireland. The data on births, marriages and deaths are collected locally by (about 300) local Registrars, while the information on adoptions is supplied to the GRO by the Adoption Board. It is intended that divorces and civil nullities (but not separations and church nullities) will be included on the new database. The data will be captured and transferred electronically by the Courts Service to the GRO.

The recording of vital events includes the capture of relevant (i.e. in the context of births, marriages, deaths, etc.) categorisation information relating to individuals e.g. address details, age/date of birth, gender, marital status, occupation (where the event is a marriage or death; in the case of a birth the occupations of both parents are recorded) and nationality (of parents in the case of a birth event).

The births, marriages and deaths information is currently recorded in paper form and is transferred within the specified time limits to the GRO. Data, anonymised in some cases, may then be passed on to other bodies (e.g. the ESRI – *see Section 4.2.3*, CSO, DSFA, Passport Office, Directors of Community Care, Hospitals, Gardaí) in compliance with relevant legislation. Compilation of vital statistics (births, deaths and marriages) by CSO is one of the major uses of the information. However, a particular concern to the CSO has been the non-availability of the national Marriage Register since 1996 resulting in the absence of detailed marriage statistics in Ireland since then.

The relevant details from the GRO paper records are electronically captured by the CSO and the results are published quarterly and annually in accordance with the provisions of the Vital Statistics and Births, Marriages and Deaths Act, 1952. In addition, and under the requirements of the Births, Deaths, and Marriages Registration Act, 1972, the CSO provides the DHC with an anonymised computerised file of the GRO data.

Statistical quality assessment

The GRO is in the process of computerising at source the information it collects– the Civil Registration Service. Apart from the information on births, marriages, deaths and legal adoptions currently collected, details of divorces and civil marriage annulments will be captured from information provided by the Courts Service. The fact that separations and church nullities will not be included renders the information to some extent incomplete in a statistical sense as the variable 'marital status' cannot be fully determined from the GRO information. As it stands, the GRO data can only be categorised to the top level of the CSO's classification of marital status (Single, Married, Widowed, Divorced, Unknown); the more detailed sub-categories under 'married' (to show separations and remarriage) cannot be

obtained at present. The PPS Number will be recorded in all cases (where it exists). At present, where the PPS Number is available, the GRO information can be linked to DSFA information.

4.2.2 General Medical Services Payments Board database for primary care reimbursement

The General Medical Services (GMS) Scheme is concerned with the provision of free general medical services for all persons 70 years old and over and for persons under 70 years who are unable to obtain general practitioner medical and surgical services without undue hardship (i.e. Medical Card holders). The Drugs Payment Scheme (DPS) is concerned with the reimbursement of payments in excess of €70 per month by eligible individuals or families ordinarily resident in the State who do not have a current Medical Card for their purchases of approved drugs, medicines and appliances. The GMSPB validates claims and makes payments to general practitioners, community pharmacists, optometrists/ophthalmologists and dentists operating under the GMS, DPS and other healthcare related schemes (e.g. Long-term Illness Scheme, Dental Treatment Services Scheme). In order to exercise its functions the GMSPB maintains a computerised database containing relevant information on all of its clients. Individual patient details held on the GMSPB central patient table are mainly supplied electronically in batch form on a monthly basis by the regional health boards.

The bulk of the information on the database relates to the GMS and DPS schemes. At present, details concerning around 2.5 million individuals are included in the database (GMS - 1.2 million; DPS - 1.3 million). In principle, all individuals in the State are eligible under either scheme but at present there is significant under-coverage in the DPS, as many people do not as yet avail of the Scheme. However, coverage is quickly increasing as public awareness improves of the possibility of capping family/household expenditure on drugs. The PPS Number is recorded for all individuals covered by the DPS; in the case of the GMS, about 60% of the records include a PPS Number (for all GMS cases, of course, a Medical Card Number is recorded).

The database holds individual characteristics as follows: date of birth, gender, marital status (not completed in many cases for DPS records), individual's county of residence and health board area, phone number (incomplete). The first level of the geographical coding is health board and the second level identifies county.

The GMSPB is concerned about the existence of a significant number of duplicate individual records on the system. Under present arrangements, an individual may have contact with different health boards and, consequently, may be allocated more than one medical card (i.e. resulting in different Medical Card Numbers for the same individual). Work is underway in the GMSPB in collaboration with the health boards to remove such duplicate records. The marking of 'inactive' records (e.g. as a result of deaths) is also a major focus of the GMSPB.

As part of a long-term solution to this issue, the Board has embarked on building a Central Client Eligibility Index (CCEI), which it views as a corner stone in developing a Unique Patient Identifier (UPI) Scheme. This, in turn, concerns the development of an integrated regional/national IT system and related infrastructure as one of the main components to underpin the delivery of integrated health care. The Board greatly favours the use of the PPS Number as the individual identifier. The availability of a PPS Number for all individuals would, in the Board's view, facilitate the desired improvement in the integrity of the database information and, in practice, allow for the increased linking of GMSPB data with information available in other relevant authorities. A current drawback, however, is that many individuals, particularly in the older age groups, do not have a PPS Number. Nevertheless, over time this situation will improve and, in principle, all individuals in the State will eventually have a PPS

Number allocated. The use of the PPS Number for linking diverse pieces of information held in diverse ways on particular individuals and for sharing this information is not currently provided for in legislation. Assuming that any technical or legal barriers can be surmounted, the greater use of the PPS Number would facilitate the expanded statistical potential of the GMSPB data. While the GMSPB uses the PPS Number where it is available, in the meantime and in the absence of a PPS Number in all cases, the payments system must continue to be administered on an ongoing basis. To do this, the CCEI system automatically allocates an identification number to each individual. In addition, a Household Number is allocated so that persons within the household can be linked for DPS payment purposes. Over time, it is expected that the PPS Number will replace the CCEI number. Currently, where a PPS Number is not available, a Medical Card Number or a Medical Services Number, in addition to address and date of birth data have to be used to identify individuals and to facilitate the removal of duplicate records.

The Board publishes an Annual Report giving details of payments made and information on developments.

Statistical quality assessment

The details held by individual health boards are more extensive but these are not passed on to the GMSPB currently as the information is not critical to administering and processing the necessary payments. In addition, the range and detail of information at health board level can vary significantly, as can the coding and classification of the data.

4.2.3 National Perinatal Reporting System

The National Perinatal Reporting System is operated and maintained by the ESRI on behalf of the DHC. Its purpose is to provide national annual statistics on perinatal events (e.g. pregnancy outcomes, perinatal mortality) with a variety of analyses of the results. The coverage is of all live births and late foetal deaths (stillbirths). Data collection began in the late 1970s and a complete database only became established in 1984. In 1999 – the latest year for which data have been published – over 54,000 births were recorded in the system.

The source of the basic data is GRO births registration forms (there are four slightly different versions of the form that are sent to the following: GRO; Director of Public Health & Medicine; ESRI (Perinatal System); Hospitals). As such, the National Perinatal Reporting system is part of the birth registration system which ensures 100% coverage but has no link with registration files on other data sets.

The Birth Notification Form (Part 3) is returned to the ESRI for coding and electronic data capture. A range of data is collected about the birth (e.g. date of birth, gender, birth weight, gestation period, occurrence of perinatal death, cause of death, morbidity, length of stay, type of antenatal care, type of feeding, whether admission was booked, maternal immunity to rubella, method of delivery, administration of BCG). Information is also captured concerning the mother and father (including date of birth, age, place of residence of parents (only county code) and their occupations. In addition, further specific information on the mother is collected (marital status, number of mother's previous live births, children still living (including children in foster care, adopted out), stillbirths and spontaneous abortions. The collection of nationality of mother and father has been introduced during 2003; nationality coding has yet to be decided. In the 1999 results, occupations of both mother and father are coded and grouped, with minor modifications, according to the CSO system of socio-economic groupings as used in the 1991 Census of Population. The 9th Revision of the International Classification of Diseases is used for disease coding. Following capture by the ESRI, the information is validated using a combination of manual and automated checking. Procedures are in place to ensure that there are no duplicate records held.

The published results (see *Report on Perinatal Statistics for 1999* published by the ESRI in association with the DHC) are used by clinicians, epidemiologists, health administrators, policy-makers, politicians, researchers and the media.

Statistical quality assessment

The database holds a considerable amount of useful data, is continually being updated for new births and also uses some standard classifications. Data for 1999 is the latest available essentially because the system is paper based. Modernisation of the registration system by the General Register Office (GRO – *see Section 4.2.1*) should improve the timeliness of the results as the information will be electronically captured and transferred by the hospitals.

4.2.4 EUROCAT (ERHA Register of Congenital Anomalies)

Surveillance of congenital anomalies in the ERHA region, with a population of 1.4 million people and approximately 20,000 births annually, is undertaken by the Dublin EUROCAT Registry (European Registration of Congenital Anomalies). The Dublin Registry, operated and maintained by the Department of Public Health of the Eastern Regional Health Authority, monitors trends in the prevalence of birth defects in the eastern region of Ireland. The Registry was established in 1979 by the Medico-Social Research Board as one of two pilot registries that led to the formation of EUROCAT, a European network of almost 30 population based congenital anomaly registers, the central registry for which is based at the University of Ulster, Newtownabbey, Co. Antrim (http://www.eurocat.ulster.ac.uk/). Multiple sources of case ascertainment are used for adding new records to the Registry including birth notifications, data from the Hospital In-patient Enquiry System (HIPE), death certificates, pathology reports, and the Domiciliary Care Allowance Scheme. Information collected through child developmental screening clinics and Public Health Nurses are other sources of notification. The data are computerised and retained in encrypted form; access is restricted to a nominated account, protected by password. Only births to mothers resident in the ERHA region at the time of birth are included. Data are collected from birth to five years of age as, while some anomalies may be obvious at birth such as spina bifida, others do not become apparent until early childhood, for example, some forms of congenital heart disease.

The following variables are routinely collected for each case: birthdate, gender, birthweight, length of gestation, type of birth, presence of congenital malformations, mother's and father's birthdate, marital status of mother, an ED variable to record area of residence of mother and a family status variable based on previous live births. The denominator populations, consisting of births to mothers resident in the ERHA region during a particular period, are sourced from vital statistics data compiled by the Central Statistics Office. The register contains a unique reference number assigned to the birth notification form, the intent of which is the avoidance of possible duplicate entries. Coding is provided by EUROCAT.

The objectives of the Registry are:

- To provide baseline epidemiological information on congenital anomalies in the Eastern Region;
- To detect and investigate trends in the frequency of congenital anomalies in order to assess the impact of known or suspected risk factors;
- To evaluate the effectiveness and efficiency of health services (primary prevention and treatment);
- To provide a well-documented database for etiologic and clinical research; and

- To act as an information centre that can respond to specific needs, such as the assessment of the impact of environmental accidents or change, or the suspicion of teratogenic influences from food, drugs or other exposures.

Statistical quality assessment

Data on approximately 13,000 cases relating to the ERHA area are captured on the register. Data obtained are from a wide variety of sources and are collected in a standardised form. Although the focus is to ascertain cases in their first instance, data is updated when possible. As data is collected using EUROCAT coding structures, it already complies with European Standardised classifications and can be used in the compilation of statistics at a European level. The Dublin Registry is an active member of the EUROCAT international equivalent - the International Clearinghouse for Birth Defect Monitoring Systems (ICDS).

4.2.5 Hospital Waiting Lists

This information on the numbers of persons waiting over three months for public in-patient and day-case hospital procedures in public hospitals is compiled by the DHC from quarterly health board returns of aggregated data supplied by hospitals in the individual health board areas. This information, compiled in its present format since 1993, gives aggregate numbers of people on hospital waiting lists for each acute hospital categorised by: broad age groups (child, adult); speciality required (e.g. ophthalmology, orthopaedic, general surgery); and by length of time waiting i.e. from the time the medical consultant makes the appointment for a hospital procedure to the end of the reporting reference quarter. Currently, the results are published showing two waiting time bands: 3-12 months and over 12 months. Apart from the broad child/adult categories and health board breakdown, no other demographic or socio-economic classifications are used.

The end-December 2002 results were published by the DHC on its website in May 2003.

Given the proposals on restructuring the national health services, responsibility for collecting and publishing information on hospital procedure waiting lists may be transferred from the DHC to a different body.

Statistical quality assessment

The data provides very important broad indicative information to policy-makers, health service providers, researchers and the general public. However, the fact that it is compiled in aggregate form, and without the possibility of detailed analysis by age, sex, area of residence or socio-economic grouping, limits the extended statistical usefulness of the existing data for broader social analysis. As regards the 'age' variable, the current child/adult categorisation is based on a child being *usually* defined as under 14 years of age. Collection of a 'date of birth' variable and an analysis of the results by an appropriate age band category would greatly improve the data. A further significant gap in the data is the lack of information on persons waiting up to 3 months for hospital procedures. The addition of this category would enhance the data and complete the picture for the 'length of time waiting' variable. Moreover, the inclusion of average waiting time would facilitate a better understanding and interpretation of the data, particularly where demographic movements can affect the absolute numbers of persons shown. In terms of current data availability and dissemination, the DHC's intention to publish a more detailed waiting time breakdown for adults and children (i.e. 3-6, 6-12, 12-24, over 24 months) is to be welcomed.

From the viewpoint of the public's perception of the data, it may be worthwhile for the DHC to review the existing definition of waiting time and to consider extending it so that it covers

the entire period of time from referral by the General Practitioner to hospital/consultant. The possibility of the HIPE system collecting information from the patient on admission to hospital on the length of time waiting for admission is suggested in the next section.

4.2.6 Hospital In-Patient Enquiry

The purpose of the Hospital In-Patient Enquiry (HIPE) database for which data are collected by the ESRI on behalf of the Department of Health and Children, is to record in-patient and day-case activity in acute hospitals. Individual information is collected and reports[21] are prepared by the ESRI giving aggregate statistical analyses of the results e.g. numbers of people discharged from publicly funded acute hospitals (including day-cases), length of stay, diagnoses, procedures performed, etc. Statistical information from HIPE is also supplied to the EU Statistical Office (Eurostat), the World Health Organisation and OECD. The HIPE data are of interest to clinicians, epidemiologists, health administrators, policy-makers, politicians, researchers and the media.

Data on HIPE relate to all discharges from both publicly funded (including voluntary) acute hospitals and two private acute hospitals that voluntarily report to the HIPE system. At present, over 950,000 cases are reported annually. This represents about 95% of cases in these hospitals. Reporting by private hospitals is expected to improve with time. The information provided to DHC-ESRI is linked to Patient Administration systems within hospitals. The vast bulk of the data is transferred electronically to the ESRI on a monthly basis as most hospitals have a windows-based HIPE data collection system operating. While individual identification information is available in the case of the Medical Record Number assigned by the hospital (and may be available in the case of the Medical Card Number), the PPS Number is not collected.

Given that HIPE is not a patient-based database, duplicate information (e.g. personal details/demographics) is collected on an individual if they present at a different hospital. A Unique Medical Record Number is assigned to each individual on admission to a hospital but this number is only unique to that hospital and is not used by other hospitals. A GMS Number is collected where available but not all people admitted to hospital have a medical card. Details of date of birth, gender, marital status, area of residence (basically county with extra detail for city areas) are also collected. Nationality data are not collected but for short-stay visitors to Ireland information is available on their normal country of residence. A pilot project has been initiated to investigate the feasibility of collecting information on ethnicity (including Travellers and other ethnic groups). The International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) developed and used in the U.S. is used to code diagnoses and procedures. Otherwise, standard classifications do not appear to be used.

Statistical quality assessment

This data source has excellent coverage and provides very useful information on the interactions of individuals with the publicly funded hospital system after they are admitted. Currently there is no obligation on private hospitals to submit data; their inclusion in the database would be desirable and it is expected that over time, a growing number will submit reports.

The major statistical consideration with this database as it currently stands is that it is discharge-based and not patient-based. Hence, it would be difficult to track events relating to an individual.

[21] HIPE ten year report (1990-1999), Public Health Information System; Casemix statistics.

On the specific question of capturing race/ethnicity data in the future, if a decision is taken to collect this information, the development of a classification on race/ethnicity that can be used for statistical purposes needs to be addressed. In this context, a co-ordination role for the Equality Authority in the development of such a classification has been discussed and agreed in principle with the CSO.

In addition, in order to connect with the information available from the Hospitals Waiting List Survey (see Section 4.2.5), the possibility should be investigated of collecting data from the patient, at the time of admission to hospital, on the length of time waiting from the time of initial referral by the General Practitioner to the hospital/consultant to the time of admission.

4.2.7 National Cancer Registry of Ireland

The National Cancer Registry of Ireland (NCRI) database records cancer incident cases in the State. Established in 1991, data collection commenced in 1994 and the registry is administered by the National Cancer Registry Board. About 20,000 cases are recorded annually. The NCRI collects data continuously from different health settings such as pathology, GPs, etc. Information is actively sought by Tumour Registration Officers based in hospitals and the data are captured from hospital administration sources, medical records, and from laboratories. Data from the DHC-ESRI HIPE (using Medical Record Number) and from deaths information from the GRO is also obtained by a mixture of manual and automated linkages. In addition, a small element of the data is obtained from GPs. Trained cancer registry nurses gather much of the morbidity data. Information on deaths is currently captured manually from GRO records. The computerisation of GRO data will improve the capture and timeliness of NCRI data.

The main focus of the Registry is on the occurrence of a cancer case, not the individual involved, although information on individuals is of significant importance. All records are considered relevant and active even after patient death. New cancers for the same person are entered and personal details are updated at the same time. Apart from details of the tumour and treatment, the NCRI captures information on age, gender, marital status, occupation, address (ED coding is being undertaken). Nationality information is not collected and cancers are only registered for Irish residents (based on address of residence in Ireland). There is provision for recording the PPS Number but in general it is not currently collected. The intention is to do so in future if the PPS Number is permitted to be used generally. At present, each registered individual is assigned a unique personal identification number by the NCRI. This allows multiple cancers in the same person to be linked over time. As part of the data capture, the Medical Record Number and the GMS Number are recorded (if available). Information is missing for some fields (if not in medical record) and occupation is only entered in about 50% of cases (particularly older/retired people). The United Kingdom Office for National Statistics (ONS) classification is used to code occupation (this is consistent with the CSO's) and the WHO's International Classification of Diseases is used to classify cancers and treatments. New treatment data are entered up to six months after diagnosis. Records are also updated in the case of death (with cause of death entered). Duplicate information is removed at source where identified. Manual and automated procedures are operated on an ongoing basis to identify and remove such duplicates. In addition, manual and automated checking is undertaken for inter-item consistency.

Overall, the NCRI covers about 95% of cancer cases and there are ongoing efforts to have exhaustive coverage. At the time of examination of the database (July 2003), about 97% of the 2001 data and about 85% of the 2002 data had been captured. The NCRI publishes an Annual Report on incidence, mortality, treatment and survival data on cancer. This information is used by health professionals and managers, researchers and the public generally.

Statistical quality assessment

The NCRI data source is currently very useful for collecting and disseminating data on cancer incidences and treatments. There are ongoing arrangements to maintain and improve the integrity of the data. The permanency of the database and the fact that it is continually being updated should tend to improve what appears to be already good quality data, In addition, the fact that the occupation coding used by the NCRI is consistent with CSO's coding further enhances the statistical potential of the data. Use of GeoDirectory information from An Post/Ordnance Survey of Ireland could be investigated as a means of optimising geographical analysis of the data.

4.2.8 National Breast Screening Programme

Breast Check, the National Breast Screening Programme was set up to screen women in the target population (50 to 64 years) for breast cancer every two years. Its aim is to reduce mortality from breast cancer of the women screened by 20% in ten years. The programme is available by invitation to all women in the age cohort by Electoral Division (ED) within county area and is free of charge. Phase 1 of the programme began screening women in the areas covered by the ERHA, the NEHB and the MHB in February 2000 and targeted over 60,000 eligible women to the end of December 2001. Almost three-quarters of those eligible presented for screening. Phase 2 will extend screening to the whole country, covering over 290,000 eligible women and is expected to begin in 2005.

The list of eligible women was compiled from data supplied by DSFA, the GMSPB, VHI and BUPA – each of which provides two files every month concerning (a) details of new women entering the programme and (b) updated information on existing women in the scheme. Procedures are in place to identify and remove any duplicate records - new records are uploaded onto the computerised database. In addition, information regarding deaths from breast cancer is supplied by the National Cancer Registry of Ireland (NCRI) and there is also a facility for self-registration by candidates. Records are continually updated from information received from all of these sources and from death notices in the press. 'Inactive' records are marked 'excluded', 'deceased', 'over age', etc. but are not permanently deleted.

The PPS Number is not currently used but the intention is to collect it if legally permitted to do so. At present, a unique 6-digit identification number is automatically generated by the computer system. Address information is not coded, but an ED code is allocated. In addition, the telephone Area Code is captured. Disability is only captured when the information is supplied by the client but it is not specifically categorised except as special needs. NBSP systems have the capacity to capture the nationality variable but it is not captured currently. There is little or no use of classifications or coding systems.

Monthly statistical reports are produced one month in arrears and more detailed information is provided six months in arrears. An annual report was published for 2001-2002 containing statistics from commencement of the BreastCheck programme in February 2000 to end 2001. The main users of the data are policy-makers, service providers, public health specialists, epidemiologists within the programme and the public.

Statistical quality assessment

As it stands and given the present incomplete coverage of the basic information sources, the NBSB is aware that the programme is covering about 92% of the targeted population. Ongoing efforts are being made by the NBSB to address this element of under-coverage, especially through promotion of self registration, and it is expected that complete coverage will be achieved over time. In addition, the present 75% take up by the public is expected to

increase over time as awareness of the programme increases through promotion, advertising, etc.

This is a potentially useful data source and should become even more useful as coverage increases and with the planned extension to nation-wide screening in 2005.

4.2.9 National Intellectual Disability Database

The National Intellectual Disability Database is operated and maintained by the Disability Databases Division of the Health Research Board (HRB). The database contains information on specialised health services either currently used or needed by people with intellectual disability. It contains information on degrees of intellectual disability of the individuals concerned but (deliberately) it does not contain diagnostic information. Established in 1995, the database informs the strategic planning for intellectual disability services. It also acts as a monitor of demographic change in the population of persons with intellectual disability, and facilitates relevant and approved research and audit.

The intention is that the database should cover all persons with moderate, severe, and profound intellectual disability and those with mild intellectual disability accessing or requiring specialised health services. Participation in the database is voluntary (informed consent is obtained) and some 27,000 people are currently registered. Reports of non-consent from health boards (the data collection co-ordinators) are very low.

The process of gathering the information into the database occurs in three-stages. Firstly, a National Planning Data Form (NPDF) which may be a paper or electronic record is completed/updated by a Service Provider (SP) twice yearly for each registered individual. The SP may be a health board, a non-statutory/voluntary body, a person responsible for a Community Care Area or a school principal. The second stage sees Regional database co-ordinators in each health board receiving the information gathered twice yearly by the designated service providers in their region. While a twice-yearly update is recommended some health boards choose to implement just one update each year. The regional co-ordinators collate the information and check the integrity of the data. If it is not already the case, all records are computerised at this stage. The data collected at this point of the process form a 'Regional Intellectual Disability Database'. At the third stage, the national database is updated using electronic returns (excluding personal data) from the seven regional health boards and the Eastern Regional Health Authority. Health Boards make returns to HRB once a year.

To facilitate data integrity, each individual on initial registration is assigned a unique six-digit Personal Identification Number (PIN) which is retained even if the individual moves between service provider or health board. The first digit of the number indicates health board in which first registration occurred. Procedures are in place to ensure that all potential new registrations are verified by the Regional Database Co-ordinator before any new PIN is allocated.

The following administrative variable categories are among those captured on the NPDF: age, gender, health board and community care area codes, as well as Planning Area and ED codes (to denote place of residence) and county code. In addition, separate variables are included for health boards with responsibility for (a) funding current services and (b) funding future services. Details of whether registered person is of no fixed abode are also captured. Optional data items include a planning area code (developed by individual health boards and therefore not unique), ED code and phone numbers of next of kin. Next of kin details are not incorporated in the HRB database. The World Health Organisation (WHO) International Classification of Diseases is used to classify the degree of intellectual disability.

National intellectual disability statistics are published by HRB in an Annual Report of the National Intellectual Disability Database Committee. The 2001 Annual Report is the latest available. Ad hoc reports for service planners primarily in DHC are also produced. The primary users of the data are health administrators, policy-makers, service planners in DHC, health boards, non-statutory bodies, lobby groups and approved researchers.

Statistical quality assessment

The database was set up solely to capture administrative data to determine current and future intellectual disability service requirements and, as such, does not capture information on specific intellectual disabilities, socio-economic or demographic details (other than area of residence, age and gender). These variables, if they were to be collected in the future, may broaden the usefulness and potential of the data source. At present, as rates of consent to participate in the database are high, the database is probably quite complete. Furthermore, arrangements to maintain the integrity of the data appear to be quite robust. The use of the PIN and computerised edits reduce the impact of duplicate data. Data validation checks are also conducted. Timeliness of supply of data has improved significantly in 2003 with all health boards submitting their data within two weeks of the required deadline.

4.2.10 National Psychiatric In-Patient Reporting System (NPIPRS)

This database, which was initiated in 1963, is administered by the Health Research Board (HRB). It provides national and regional information on in-patient psychiatric services and on the users of these services. Cross-sectional data are presented nationally, regionally (by health board area), locally (by individual hospital) and by hospital type. The information produced has been used for planning, research and policy purposes in the mental health area since 1965. The database captures the total psychiatric (hospital and unit) in-patient population in terms of admissions and discharges (as well as deaths). It is therefore event based and, in the context of its principal purpose referred to above, it is not especially concerned with the linking of events associated with individuals or with the compilation of person-based statistics. The database structure is also used in carrying out decennial censuses of the in-patient population. Thus, as the database is event-based and not person-based, a person may have several admissions during the course of the year, each admission being recorded separately.

All psychiatric hospitals, units, private hospitals and children's centres report to the HRB on an annual basis (quarterly from 2003) giving details of admissions and discharges at an individual level. Returns are made in both paper and electronic form. In addition to diagnosis information captured on admission and discharge (classified according to the WHO's International Classification of Diseases), details are collected on age, gender, marital status, address (health board area, townland, county), occupation (using CSO Census of Population 1996 classification), and employment status (unemployed/employed, retired, student, house duties). In addition, legal category of admission (voluntary, non-voluntary) and order of admission (first ever, other), along with discharge reason and length of stay are recorded. All fields are mandatory and true missing values are recorded as such.

In 2001, there were 24,446 admissions (of which 7,301 were new admissions) and 24,304 discharges (including 266 deaths). At any point in time there are about 4,000 in-patients residing in these hospitals.

Results are published annually (Activities of Irish Psychiatric Services) with annual regional bulletins for each health board area. Relevant database information has been included on the Public Health Information System from 2003. The information is used by clinicians, epidemiologists, health administrators, policy-makers, researchers, the general public along with various mental health advocacy groups.

Statistical quality assessment

This long-established database holds very useful data on admissions and discharges and some key categorisation details are captured. It provides a comprehensive national picture of provision, use and users of the psychiatric in-patient services in Ireland. Regular validation of the data is carried out throughout the process and finally before publication. In addition, once an individual is discharged his/her record relating to the event recorded is not carried forward (but it is archived). As mentioned, the database records are event based, and because there is no unique personal identifier used, certain categorisation data relating to individuals can understandably be duplicated (or possibly even recorded differently on different occasions) on the database under different event records (e.g. if the same person was admitted to different hospitals and was allocated different local patient numbers). Hence, and while acknowledging that the database was not established to do so, it is not readily possible to track data relating to an individual in a current or in a historical sense. As regards timeliness, there is about a one year timelag in the publication of results but this will improve with the move to quarterly collection.

4.2.11 Notifiable Infectious Diseases Database

The National Disease Surveillance Centre (NDSC) maintains reporting systems for notifiable and non-notifiable infectious diseases, antimicrobial resistance, immunisation and infectious disease outbreaks.

These systems were introduced in the late 1990s and are concerned with recording statutorily notifiable disease events and other data (as reported by GPs, hospitals or clinical testing laboratories and, in a related way, all persons in the State presenting with notifiable infectious diseases). The Department of Health has been responsible for monitoring the incidence of infectious diseases since 1947 and data collection on this subject therefore predates the establishment of NDSC. The 2001 Annual Infectious Diseases Statistical report contains time series of data from 1982 to 2001, although reporting to DHC and NDSC prior to July 2000 was in aggregate form.

At present, the NDSC data is abstracted from information captured locally (health board area). Apart from data specific to the disease reported, the current minimum dataset comprises: health board identifier, disease, date of onset, date of diagnosis, gender, date of birth, age, country of birth, vaccination status, laboratory confirmed, organism and date of notification. Apart from the use of the WHO International Classification of Diseases for classifying the various diseases and the Standard Nomenclature (SnoMed) for clinical terms, standard classifications are not used for other categorisation variables. The NDSC data source cannot be linked directly to any register system.

Currently, it is thought that only a small proportion of all notifiable infectious disease cases are reported by GPs, hospitals etc. In the case of serious diseases, however, reporting is thought to be virtually complete. Despite the significant under-reporting of infectious disease it is felt that this is reasonably constant, allowing analysis of trends. Nonetheless, the low reporting level must have negative implications for enhancing the statistical potential of the data as it currently stands. It is hoped that revised infectious disease regulations will improve this situation. In addition, the impact of duplicate individual data is a concern to the NDSC and work is ongoing to remove duplicate and inactive records.

A new Computerised Infectious Disease Reporting (CIDR) database system is being developed by NDSC with a view to implementation in 2004. It is intended that the relevant information from clinical notifications and from laboratories will be captured electronically at local level and linked appropriately. The new system should facilitate faster publication of the

results and the intention is to move as closely as possible to a 'real time' data-reporting basis. Apart from the information referred to above, additional data will be collected in relevant disease specific cases (e.g. nationality, race/ethnicity, membership of Traveller community, occupation). Geographical assignment to an individual case i.e. by county, community care area or Electoral Division will be possible in the future. If use of the PPS Number is permitted then the integrity of the data will improve as the possibility of duplicate information for individual persons (as distinct from disease events) should be greatly reduced.

The NDSC publishes weekly (quarterly in some instances) and annual data on reported cases of notifiable infectious diseases with breakdowns by age, gender and health board area. Quarterly data for Sexually Transmitted Infections are also published. The annual 2002 data are not available as yet (but weekly data are published on the NDSC's website. As it stands, the NDSC data is used by DHC (policy, planning), researchers, public health departments, health boards as well as environmental and agricultural and food processing interests.

Statistical quality assessment

Very useful data are available from this database, particularly for the more serious notifiable infectious diseases for which reporting is excellent. For the less serious diseases, the low current level of reporting probably limits the expanded statistical use of the data. In addition, the collection of only a limited amount of demographic and socio-economic data i.e. only for certain diseases, further constrains the expanded statistical quality of the data. However, the introduction of the new decentralised computer system (CIDR) should greatly help in removing some of these shortcomings.

4.2.12 Annual Survey of Long Stay Units

DHC collects aggregate data on activities of over 550 long-stay units[22] (i.e. public and private nursing homes) in the various health board areas in order to compile statistics on the number of beds available for long term care, how the beds are used and the types of patients occupying them. The information collected by the DHC covers the number of patients classified by age (in age band), gender, health board area as well as level of dependency, socio-medical status (chronic mental illness, chronic physical illness, mental infirmity/dementia, mental handicap, social reasons, physical disability, terminal illness, convalescence or rehabilitation, other reasons), source of admission, destination on discharge and length of stay. No individual person data are returned to the DHC.

The results are published annually by the DHC, the latest report relating to 2001. In 2002, 495 or approximately 87% of registered nursing homes made returns for these units covering over 20,000 patients, almost two-thirds of whom are over 80 years of age.

Statistical quality assessment

Individual record data is not available to the DHC. Aggregate information is collected for each long-stay unit and the results are compiled for a very specific statistical purpose that the data serve. As such, there is very little information collected to allow further key categorisation of the aggregate data (e.g. there is no information on marital status or on socio-economic grouping). The limitations in linking aggregate data would have to be tackled before any further progress can be made in terms of linkages to other data internal or external to the DHC.

[22] The categories are: health board geriatric home/hospital, welfare home, district/community hospital, voluntary geriatric home/hospital, private nursing home.

4.2.13 Health Behaviours in School-aged Children

The Health Behaviours in School-aged Children (HBSC) survey is a cross sectional study conducted every four years on behalf of the Department of Health and Children by the Department of Health Promotion, National University of Ireland – Galway. The survey was first undertaken in Ireland in 1998 and is part of a World Health Organisation Cross-National HBSC Study which is a research study of the health behaviours and health of adolescents across a significant number of European and North American countries. The results provide baseline information on health related and social behaviours among school-going young people aged 10-17 years. The survey is of a tick box format and collects a wide range of information on health behavioural characteristics and social attitudes and relationships of young people. It is self-completed by respondents. Information is collected on age, sex, parental occupation and employment status, perceived family affluence and lifestyle and the health board area of the school attended. In addition, information on the respondent's living arrangements and relationships within the immediate and extended family is collected. No individual identifier is collected and standard classifications of characteristics are not used.

In 2002, 176 schools (out of a total of 351 contacted from the Department of Education and Science school lists) covering 8,316 pupils participated in the Survey. Of these, responses from 5,712 pupils from 93 schools were selected to maintain seasonal comparability with the data from the initial (1998) Survey. The main findings are published in the report - The National Health & Lifestyle Surveys.

Statistical quality assessment

The HBSC survey instrument is an international standard questionnaire used by all participating countries. It enables the collection of a vast amount of standardised data and thus enables the quantification of patterns of key health behaviours, health indicators and contextual variables. These data allow cross-national comparisons to be made and, with successive surveys, trend data is gathered and may be examined at both the national and international level. The HBSC International Reports that contain the published data are useful for policy-makers as they can compare the position of their country in relation to others not only on health behaviours and health outcome measures but also on health determinants. The information, however, is more qualitative than quantitative and relates to young people's perception of their living conditions, behaviour and lifestyles.

4.2.14 Survey of lifestyles, attitudes and nutrition (SLÁN)

This survey was introduced in 1998 and is conducted at four-yearly intervals. Its purpose is to provide baseline information on health related behaviours among adults 18 years and over. It is closely related in form and content to the Survey on Health Behaviour in School-aged Children and is conducted on behalf of the DHC by the Department of Health Promotion, NUI Galway and the Department of Epidemiology and Public Health, UCD. Potential participants are selected from the electoral register. In 1998, 10,515 were covered by the sample while 11,212 were sampled in 2002. Response reached 5,992 (or 53%) in 2002. The survey is designed in a tick box format and is generally seeking qualitative data although some quantitative information is requested. The survey collects a range of information on individuals' lifestyles, attitudes, health status, nutrition, relationships, etc. The following categorisation variables are captured: age, gender, marital status, occupation, education, categories of household net income, the ED and County of residence, long-term illness/disability, nationality (as per Census 2002 question), membership of religious grouping, farm/non-farm household, holder of Medical Card and car ownership. A unique personal identifier is not used. Standard classifications are generally not used, although

where possible, recommendations from the EU health monitoring programme have been incorporated.

The latest 2002 results are published in the report - The National Health & Lifestyle Surveys. The results are used by a range of stakeholders including policy-makers, service providers, public health specialists and researchers.

Statistical quality assessment

This is a cross-sectional point-in-time survey conducted at four-yearly intervals which provides very valuable qualitative information on the behaviours, and perceptions of people in the general context of overall health and well-being. While quite a number of key categorisation variables are collected (in tick box format) there is little use of standard classifications and coding (apart from the use of recommendations from the EU health-monitoring programme where possible). Nor is the PPS Number (or other type of unique identifier) used.

4.3 Other data sources

There was not sufficient time in this project to examine in detail the remaining data sources identified by the SGSES. However, it is probable that the same issues would emerge, and similar conclusions be drawn, as for the sources examined.

4.4 Statistical potential

All of the data sources examined were fulfilling already specified statistical purposes and a number were also addressing administrative needs. While it was clear that there are some linkages between some data sources, these linkages were primarily designed to assist in database establishment (e.g. in identifying clients) or in facilitating basic data input and improving data quality. The team did not find any instance in the sources examined where statistical outputs were linked across databases across the ambit of the DHC. Such possibilities would greatly enhance the existing statistical potential of the data sources.

To realise this potential within the health domain and assuming the appropriate legislative arrangements were put in place, it would be necessary to adopt the use of a unique personal identification number (such as the PPS Number) in all the data sources examined. This would potentially enable information on individuals to be linked at individual record level from all health-related data sources and for a greatly expanded range of aggregate statistical outputs to be produced. However, to enable meaningful analysis, the demographic and socio-economic variables of interest (e.g. date of birth, area of residence - or other geographical variable, marital status, occupation, nationality, ethnicity and socio-economic grouping) would also have to be available using recognised commonly used classifications and coding systems. In addition, the use of standardised definitions and concepts in capturing data, as well as the education of all involved on the importance of the data's wider statistical information context, would be needed 'across the board' to underpin accurate data capture and improved results compilation. The possibility of being able to produce a sound statistical health profile of the population from birth to death seems very attractive and, if realisable, would be extremely useful for guiding policy-making in the health area. Apart from point in time statistics, an integrated framework of health data would enable longitudinal analyses of the population from the point of view of their health status and their interactions with health services over time and under changing conditions (e.g. if residence or occupation changed).

Apart from purely health related statistics, the use of the PPS Number for linking individual health data with data held in other government departments could lead to an even more expanded statistical picture of the population in the context of health, education, housing, taxation, poverty, etc. and the way these factors impact on the different social dimensions of the population. A much strengthened body of information for policy and other needs would result.

4.5 Overall conclusions and recommendations

4.5.1 Conclusions

The field of health statistics is wide ranging, the various data sources being developed over the years to address both administrative and statistical needs as defined by the DHC. The data sources examined were primarily chosen for the following reasons:

- Where it was felt there may be scope to enhance and/or improve already existing national administrative data systems (e.g. GMSPB, GRO);
- Where the data sources provide significant information concerning the state of general health of the population or subsets thereof (e.g. NCRI, HBSC, BreastCheck, NDSC, Perinatal Data, SLÁN Survey, etc.); and
- Where the data sources provide information on the interactions between the population requiring health services and the provision of these services (HIPE, Hospital Waiting Lists, Survey of Long-Stay Units, NPIPRS, National Intellectual Disability Database).

In examining these sources, the team generally concluded that while they essentially address the specific statistical requirements they were designed to meet within DHC (and apart from their administrative purposes), they largely exist as isolated islands of information which cannot currently be integrated. If such integration became possible for statistical purposes at least e.g. through a legally permitted use of the PPS Number for linking anonymised individual records, then a vastly expanded statistical potential could result from integrating these sources not only internally within the DHC but also on a cross-departmental basis (e.g. with DES, DSFA, DoEHLG and Revenue data). A comprehensive statistical profile of the population (or subsets thereof) could then be produced for both cross-sectional and longitudinal analyses.

The health domain data needs, identified as high priority and requiring development in departmental submissions to the SGSES, may not in some cases have been fully elaborated. It was clear to the team, however, that data relevant to some needs articulated in the work of the SGSES are available from DHC sources (e.g. HBSC, SLÁN) and from CSO sources (Quarterly National Household Survey, EU Survey on Income and Living Conditions and Vital Statistics).

The DHC SPAR team's specific conclusions may be described as follows:

(i) There is a vast body of data on health service provision and health status matters in existence across all the sources examined. In most cases, the information systems were originally designed to serve specifically defined administrative, policy-making and statistical requirements. There are ongoing efforts on the part of data holders to improve the quality of the information to meet these needs.

(ii) Much of the data used is currently captured at local or regional level (e.g. health boards, hospitals, GPs). There appears to be some variation in the type of information collection and capture guidelines interpretation at local and regional level. Furthermore, in many cases data are captured and processed in a fragmented and non-standardised way. As a result, there may be: (a) variation in the range and type of information required; (b) different definitions in some cases where an attempt is being made to collect the same variables; and (c) different classifications and coding systems in use for key categorisation variables (e.g. occupation, marital status, and location of residence).

(iii) In general it was found that while there may be provision within systems for collecting particular variables, the information may not end up being captured. In addition, in certain cases, information may be captured on paper records but may not be computerised in the end. Due to time constraints in this examination, a full statistical analysis of the various data sources was not possible. However, such an exercise would be required to establish the degree of completeness of each data holding.

(iv) In some cases the information available to the DHC may be a subset or selection of the information collected at individual record level. The availability centrally of complete statistical information at an individual record level would enhance the statistical potential of the data.

(v) Computer processing systems different to those available within DHC may be used at local or regional level and this can therefore hinder the task of integrating data centrally.

(vi) The team considered that some enhancements could be made to the data in certain cases with a view to making it more useful for broader social statistical analysis. The Department's Public Health Information System brings together detailed information from a wide range of health sources to meet a number of broad aims. However, each information system remains a discrete and unlinked operation and by and large does not contain data on variables which would improve its value for social analysis. Up to now this latter aspect has been considered of secondary importance to the main purposes for which the data sources were developed. Introduction of any enhancements to data sets would inevitably involve some cost to data holders while the benefits would be spread among a larger group of users.

(vii) Maintenance of the integrity of individual health data records can be difficult because of the lack of a standard centrally generated unique personal identifier. Different data requirements may use a variety of identifiers and an individual may be allocated a second (or third) identification number by different authorities. This can result in duplicate, redundant or incorrect data being held on databases making cross comparisons and linkages with other data sources extremely difficult.

(viii) Database integrity maintenance is a concern to those departments or authorities that administer such databases (e.g. the DSFA Central Records System and the GMSPB Central Client Eligibility Index). A large amount of duplication of effort can exist across these bodies, each essentially attempting to ensure, in varying degrees of isolation, that it has accurate categorisation information captured on each individual and that it has no duplicate records. It would seem to make sense not only from a statistical viewpoint but also from an administrative one that key categorisation be captured centrally only once and that it be maintained up to date thereafter.

4.5.2 Recommendations

It has already been noted in Section 4.5.1 that the possibility of the data sources examined being used in a wider statistical context either in aggregate terms, or through linkage at individual record level, has to date been thought of as secondary to the principal use of the

data. However the DHC SPAR team considered that in certain cases some enhancements could be made to data with a view to making it more useful for broader social statistical analysis. These enhancements would materially address some of the concerns raised in the recent Brennan and Prospectus reports about the capture, maintenance and analytical capabilities of existing data. They are also broadly in line with the Government's proposals on improved information availability through the establishment of the Health Information and Quality Authority. In addition, according to the DHC, the intention of the National Health Information Strategy which is expected to be published shortly, is to ensure that the best information available should be used at all points in the health care system in order to:

(i) Enable better management of finite health resources;

(ii) Provide better information to the public; and

(iii) Deliver effective and efficient services to all.

The team is of the view that its recommendations will probably coincide broadly with those likely to emerge in the Strategy concerning information gathering, dissemination and analysis.

The following general recommendations are made by the DHC SPAR team:

(i) Where it does not exist, greater standardisation should be implemented within the DHC for paper and electronic form design and in the definition and specification of key demographic and socio-economic categorisation variables (date of birth, gender, marital status, location of residence, nationality, race/ethnicity, substance intake, occupation, socio-economic grouping, etc.) and in their related classification and coding systems. Two options can be considered for this:

(a) Capture of the relevant data locally for each DHC data source; or

(b) Capture of data once in a centrally maintained database.

The first option is probably feasible but leads to the same information being collected repeatedly and is therefore sub-optimal. Multiple capture of the same data makes it difficult to ensure the integrity of the data and is less efficient overall than the second approach. The second option involves the data being captured and, where relevant, being updated by the DHC authority responsible for the operation and administration of the database. The centralised key data capture and maintenance approach should result in a much more robust body of critical categorisation information. While initial costs could be significant, the potential savings over time from dispensing with the need for multiple non-integratable databases should also be significant.

(ii) While the previous general recommendation focussed on data management within the DHC, the SPAR team was of the view that the second option could be considered in a more general light for administrative data holdings pertaining to social statistics across a number of departments and bodies. Under this scenario, it is envisaged that an independent body would be responsible for the maintenance and administration of the centralised key individual information.

(iii) Effective and efficient implementation of the two previous recommendations would be contingent on the implementation of standardised use of and access to the PPS Number, with appropriate legal arrangements for data protection covering prescribed use of the data. Assuming the legal safeguards and assurances were to be put in place, use of the PPS Number should be facilitated for the purposes of maintaining the integrity of individualised data records held and for linking different data sources at individual record level in order to facilitate production of improved statistical

information. The use of the PPS Number, along with the use of key demographic and socio-economic information (date of birth, gender, marital status, occupation, socio-economic grouping, location of residence, nationality, race/ethnicity) and related standardised classifications and coding systems, would enable the integration of individual records from different data sources and the compilation of a greater range of analyses giving a more comprehensive statistical picture of the population of health service users. The team recognises the difficulties to date in introducing the PPS Number into the health sector particularly in those areas where there is mix of public and private activity in the provision of services. This issue was highlighted earlier in the Health Strategy (*Quality and Fairness: A Health System for You*) and also featured prominently in the 2002 Annual Report of the Data Protection Commissioner. The role and use of the PPS Number for improving the statistical potential of administrative data is addressed specifically in Chapter 8 of this present Report.

(iv) Given that 'marital status' is a key socio-demographic variable, and as the GRO data on this variable is confined to marriages, divorces and civil nullities registered within the State, it is recommended that the GRO data capture arrangements be extended to collect additional data on legal separations occurring within the State. This would facilitate the generation of more complete analysis of information on the basis of marital status.

(v) In addition and in the interest of compiling the full range of vital statistics, it is recommended that the modernisation of the GRO's system of registration of all vital events be completed as soon as possible.

(vi) In the interest of improving the information on hospital waiting lists in the short-term, it is recommended that:

(a) Average waiting times; and

(b) The category 'waiting times up to 3 months' be included in the results.

Beyond this, it is recommended that the DHC review its definition of waiting time to cover the entire period from the time of initial referral by the General Practitioner to the date of hospital admission.

(vii) In order to connect with the information from the Hospital Waiting List Survey, it is recommended that the HIPE system captures, from the patient on admission to hospital, the waiting time from the date of initial referral by the General Practitioner to the hospital/consultant to the date of admission.

(viii) The use of the GeoDirectory (maintained jointly by An Post and the Ordnance Survey of Ireland) or an equivalent facility for assigning geographical (x,y) co-ordinates to standardised address fields should be considered. The use of this type of locational attribution, along with the use of suitable mapping software, enables the generation of geographical analysis at county, ED, postal district or other defined area.

(ix) Greater encouragement and education at national, regional and local level on the wider statistical use of the data captured and the need to have complete information recorded in a standardised way.

(x) Implementation of compatible computer processing arrangements to facilitate data sharing.

(xi) Implementation of improved data dissemination arrangements concerning the frequency and timeliness of information and its analytical presentation in terms of the demographic and socio-economic classifications used. Over time, it could be envisaged that key statistical information of interest to policy-makers, planners, service providers, researchers and the public at large could and should be available on a virtually 'real time' basis.

Chapter 5

Department of Justice, Equality and Law Reform

5.1 Introduction

The main areas of responsibility of the Department of Justice, Equality and Law Reform (DJELR) are:

- Crime and security;
- The Criminal justice system (the Gardaí, the Courts and the Prisons);
- Reform of the criminal law and updating areas of the civil law;
- Immigration policy and strategy in relation to asylum seekers; and
- Policy in relation to equal treatment generally, anti-racism, disability, equality and childcare.

In the criminal justice area, the principal data holdings within the department and its agencies are:

- The Garda PULSE system, which maintains all police records;
- The Courts Case Tracking System (CCTS) which covers criminal records within the courts system; and
- The Prisoner Records Information System (PRS) which maintains all prisoner records.

The two key data holdings in the immigration area are:

- Registered non-EEA nationals; and
- The Refugee Asylum Commissioner database.

A number of other data sources, which are not of major significance from the point of view of their statistical potential, also exist. These include:

- Dedicated one-off surveys in the gender equality area on housing, parental leave and transport;
- Measures set out in the National Development Plan on equality;
- Gender equality in the membership of State Boards;
- Data holdings in respect of civil and criminal legal aid;
- Child abduction; and
- Maintenance recovery.

5.2 The Garda PULSE system

The PULSE system, which was introduced in 1999, supports the work of the Garda Síochána by enabling relevant information to be recorded, processed, stored and retrieved. PULSE is primarily an operational tool providing the Gardaí with information on the recording and subsequent investigation of reported crime, thereby servicing operational and strategic decision making and informing the efficient deployment of resources within the force.

The PULSE system also underlies the publication of annual crime statistics, which are issued as part of the report of the Garda Commissioner to the Minister of Justice, Equality and Law Reform. In July 2002, the Government established an expert group on crime statistics to examine and make recommendations on the collation and presentation of information relating to reported crime. The Group was constituted in early 2003 and had its first meeting in March 2003. It has met on a monthly basis since then.

The main terms of reference given to the Group were:

- To examine the collation of information relating to crimes reported to and recorded by the Garda Síochána, mindful of the capabilities of the PULSE system;

- Examine the 'Principal Offence' rule which is used by the Garda Síochána, which is 'incident focused' rather than 'victim focused', and other counting rules used by the Garda Síochána;

- Examine the distinctions currently made between offences classified as indictable and non-indictable, as contained in the Commissioner's Report, having regard to the seriousness of the crime committed, with a view to possible changes in the classification of offences; and

- Examine the collation of information relating to other crimes where the Garda Síochána is not the prosecuting authority, for example, 'white collar' crimes prosecuted by the Revenue Commissioners.

The expert group has effectively been mandated to realise the statistical potential of the Garda PULSE system. The work of the group was informed by a public consultation process, the results of which identified a number of shortcomings in the present method of compiling crime statistics. These include poor timeliness, absence of a satisfactory level of geographic breakdown of the published data, and the need for regular crime victimisation surveys. Other limitations, which were identified, include a lack of consistency in the published data over time, poor methodological notes, and unsatisfactory publication arrangements.

Notwithstanding the fact that the expert group has been charged with carrying out a detailed analysis of the potential of the PULSE system to produce crime statistics, the SPAR team would have wished to have been able to carry out a preliminary analysis of its own at this stage. However, given that the full record layout of the PULSE file was not made available to the SPAR team, this analysis was effectively precluded.

The expert group, in collaboration with the Garda authorities, is currently focussing on the following statistical processes used to derive crime statistics from the PULSE system: target population (i.e. the distinction between reported and recorded crime), data capture, coding rules and manuals, classifications, editing procedures, data quality assessment and presentation and publication of results (including release mechanisms).

The setting up of the expert group underscores the general level of dissatisfaction that exists among the user community in relation to the quality of the existing crime statistics. The group has a year from the date of its inauguration in which to issue its findings.

5.3 The Courts Case Tracking System

The Courts Case Tracking System (CCTS), an Oracle relational database, is in operation in the Dublin Metropolitan District Courts and Limerick District Court offices. Deployment of CCTS nationwide will commence later on in 2003.

The CCTS application provides for the processing, scheduling, enforcement and reporting of criminal cases. Each case is assigned a unique identifier. The system enables users to create cases, add and adjourn court hearings, produce summonses and warrants and copy orders and a variety of other documents that are required in the course of a criminal case.

The system provides the facility to search for information by a variety of means including defendant and prosecutor names, case number, courtroom date, etc. Another advantage is the ability to derive statistics through the use of automated queries rather than the manual methods used heretofore.

Future developments of CCTS planned for later in 2003 will allow the transfer of selected data electronically between the Courts Service, the Gardaí, and the Department of Environment and Local Government. This new feature will facilitate the introduction of the full Penalty Points system on a nation-wide basis. The CCTS database will be rolled out to all District Courts by the end of 2003.

The main focus of the CCTS has, to date, been on criminal cases prosecuted by the Gardaí. However, non-Garda offences, such as those prosecuted by An Post and the Companies Registration Office, are also catered for in the system.

The extension of the CCTS to cover criminal cases in higher courts and to embrace the wider area of civil cases is being planned at present.

The main function of the database is to support the administration of the courts system and to provide management information to ensure its smooth running. The impression formed from the available information is of a comprehensive database capable of yielding good quality data mainly of interest to agencies in the criminal justice area. The management information capabilities of the system would tend to outweigh its statistical potential.

5.4 Prisoner Record System

The Irish Prison Service uses a system called PRS, the Prisoner Record System. The PRS serves two purposes. The first of these relates to the administration of records of prisoners in custody and holds personal details such as name, age, address, socio-economic status and nationality. Each prisoner is given a unique reference number (PRIS number) which is used to track the prisoner through the system. Daily files are supplied to Garda headquarters by the Irish Prisons Service, detailing those committed and released, to allow the Gardaí to keep their records up to date.

The second purpose of the PRS is to provide data to feed into the management reports and the annual report of the Irish Prison Service. The annual report deals with developments in the Prison Service in the previous year and contains statistics on activity levels. Details relating to the number of committals by age, gender, county, nationality, type of offence and

length of service are also provided. However, where a person enters two different prisons in the one year, the PRS records this as two committals and hence there is a distinction between the number of committals each year and the average number of prisoners in custody.

It is vital from a social policy and sociological perspective to be able to examine in detail the profile of the prison population to help to throw greater light on the causes of crime. A detailed profile of the prison population by age, socio-economic status, home district and nationality is a necessary requirement of any policy analysis of crime and its causes. By establishing the profile of a typical offender committing a typical offence or crime, policy-makers can derive the factors or potential causes that influence certain people to commit crimes.

While at the overall level, the output from the PRS system is relatively comprehensive, the system itself does suffer from a number of shortcomings. Not all information on the system is mandatory and in some cases the service is dependent on the goodwill of the individual being committed to volunteer certain information. In addition, the PULSE number used by the Gardaí in recording the crime is not held on the system but is held separately on another system. Individual prisons appear to operate independently of each other when using the PRS. Therefore, the details of a prisoner who has served in another prison cannot be accessed. Notwithstanding these shortcomings, the PRS is a valuable potential source of useful statistical data.

5.5 Registered non-EEA nationals

The Garda National Immigration Bureau (GNIB) was set up in May 2000 and is responsible for all policing matters on immigration. While the immigration process was computerised at the time the GNIB was set up it did not go live until September 2001. This system keeps track of all registered non-EEA nationals[23], a process that has been in existence since 1946. The system is now live in 23 locations with other locations filling out paper forms that are subsequently added to the system. The system interfaces with information from other government departments, namely: work permits, visas and asylum seekers. Each of these three areas is updated weekly.

All non-EEA nationals who remain in the State for more than three months are required to register with their local Garda Registration Office, which is located at the nearest Garda Superintendent's Office or, in the case of those within the Dublin Metropolitan Area, with the Garda National Immigration Bureau's Registration Office, 13-14 Burgh Quay, Dublin 2. The foregoing does not apply to asylum applicants who are given a temporary residence certificate by the Refugee Applications Commissioner. During 2002, over 97,000 non-EEA nationals presented for registration in the State. This figure excludes persons under the age of 16, female spouses of Irish nationals as well as persons seeking asylum and is not a point in time stock figure.

The non-EEA nationals database is potentially a very valuable one whose coverage extends to all non-EEA immigrants. The following basic demographic variables are maintained for each person on the database: sex, date of birth, marital status, country of birth and country of nationality. Given the scope and coverage of the registration system involved, it should be possible to compile accurate stocks and flow figures for the relevant target population. This

[23] A non-EEA national is defined as a person from outside the European Economic Area (i.e. the EU, Norway, Liechtenstein and Iceland). Switzerland also has a free movement agreement with the EU but is not a member of the EEA.

would be of major importance in improving the accuracy of the annual population estimates published by the CSO.

The situation in relation to information on immigrants, held by other government departments, is less satisfactory and can best be described as patchy and disjointed at present. In a number of cases where country of birth and nationality is asked, the relevant information is not always captured and even where it is, the level of disaggregation may be minimal.

A person coming to Ireland from a visa[24] required country must apply for a visa in advance of travel. Applications are made in Irish embassies or consulates abroad. In countries where there is no representation, applications can be made to the Department of Foreign Affairs in Dublin. The embassy data has not been fully computerised and, in particular, does not distinguish details of nationality. The data is supplied about six months after the end of the relevant year. Computerisation of the issue of visas is currently under consideration by the Departments of Justice, Equality and Law Reform and Foreign Affairs. About 100,000 visas were issued in 2002 compared with 35,000 three years previously. The availability of such information on a timely basis would also add to the accuracy of migration estimates.

5.6 Asylum seekers

DJELR is responsible for asylum policy. In order to discharge this policy, DJELR requires statistics on applications, trends, flows, etc. from the various agencies. These include the Refugee Application Commissioner (RAC) and the Refugee Application Tribunal (RAT) – for appeals and the Reception Integration Agency (RIA) which looks after the direct provision of accommodation for refugee applicants.

Under the UN Convention on the treatment of refugees (1951), all persons who arrive at our borders and claim asylum must be allowed into the country. All asylum seekers are provided with accommodation and given some money for living expenses by the RIA.

Detailed statistics are available on the number of applications and numbers processed by nationality and by country. However, data retrieval presents a major problem. DJELR recently commissioned a report by a private consultant to see how IT systems can be improved to enable the information on them to be integrated. The report recommended an integrated case management system for the RAC and RAT with a system which will track a person through the whole asylum process and link to the Reception Integration Agency (RIA) for those person who are granted refugee status. Implementation of this system could cost up to €8 million.

Up to half of all refugee applicants fail to turn up for their meeting at the RAC. This may be due in part to large numbers of them staying in the State for short periods and then moving back to the UK. Once granted refugee status the applicant moves onto the register of non-EEA nationals. The backlog on the asylum seekers system is defined as anyone in the system for more then six months. At present this backlog is 1,700 persons.

In the past about half of all asylum seekers moved on to the so-called Irish born child scheme (IBC), whereby the parents of a child born in Ireland apply for residency status. However, since the Supreme Court judgement in January 2003, which found that parents of

[24] The granting of a visa allows the holder to present himself or herself at Irish immigration controls to seek entry into the State. Not all non-EEA nationals are visa required. The Aliens Amendment (Visas) Order, 2002 sets out which nationalities are exempt from the visa requirement.

an Irish born child have no automatic rights of residency, this pattern of movement onto the IBC scheme may alter.

There is general acceptance that the full statistical potential of the immigration databases is not being realised at present. In this context, the need for a dedicated statistical resource within the area has been articulated.

5.7 Other data sources

Dedicated one-off surveys

The Irish Government has adopted mainstreaming of equality as a strategy to promote equal opportunities between women and men in the National Development Plan 2000 to 2006 (NDP). Mainstreaming equality between women and men means the *'(re)organisation, improvement, development and evaluation of policy processes, so that a gender equality perspective is incorporated in all policies at all levels and at all stages, by the actors normally involved in policy-making'* (Council of Europe definition). The NDP provides funding for over 130 measures, in six operational programmes - employment and human resources; productive investment; economic and social infrastructure; Peace, and two regional programmes. Equality between women and men must be mainstreamed in all but six of the NDP measures.

In order to assist implementing departments and bodies in meeting their mainstreaming requirements, the NDP Gender Equality Unit of the Department of Justice, Equality and Law Reform was set up in 2000, to provide advice, training, information and statistics relevant to gender equality to these bodies. Research was commissioned on the availability of existing gender-disaggregated statistics. This pointed out areas where there was a dearth of data.

Arising from this research, one-off surveys were carried out in the Housing and Transport sectors - two areas with poor statistical provision. The results were published in two reports.

Research was also commissioned to obtain reliable information on the uptake levels of parental and *force majeure* leave. A postal survey of some 655 employers accounting for 67,182 employees was carried out in July/August 2001. The main findings were published in a report dated 29 April 2002[25].

Equality measures in NDP

Statistics are provided on projects carried out under the Equality for Women measure of the NDP. These include aggregate funding and number of projects classified by five key strands and by NUTS2 Region. The 5 key strands are:

- Access to employment, education and training, with a particular emphasis on retraining and upskilling of women;
- Career development for women;
- Entrepreneurship among women;
- Innovative projects for disadvantaged women aged over 50 years of age; and
- Promoting gender balance in decision making.

[25] See Report of the Working Group on the Review of the Parental Leave Act 1998 on DJELR website.

This information can be considered more as management type information rather than official statistics.

Gender equality in the membership of State Boards

Basic data is maintained on the number of women on State Boards and on the percentage who are chairpersons of State Boards. Details on total board membership, and the proportion of which is appointed by the Government, are also maintained. Clearly this is an important measure within the gender equality area. However, other than satisfying the immediate policy needs, there appears to be no further potential from a statistical perspective.

Data holdings in respect of civil and criminal legal aid

Data relating to *Criminal Legal Aid* deals with claims for payments made by senior and junior counsel arising from legal services provided in criminal cases where defendants do not have the means to pay their own legal costs. In this system, the barrister making the claim is the client rather than the person who is availing of the service. Financial and activity data detailing the number of claims processed with reference to the type of court and by barrister are produced on a monthly basis and circulated within DJELR. There is no annual report or published statistical data.

The CSO does not see any potential for further development of this data source, which serves principally as a management and financial reporting facility.

The Legal Aid Board provides legal aid and advice in *civil* cases to persons who satisfy the requirements of the Civil Legal Aid Act, 1995. The primary requirements for assistance under the scheme are based on a means tested income threshold and a requirement that there must be merit in the case concerned. The Board supplies the services of solicitors and where necessary barristers through 30 local law centres situated around the country. In addition the Board has responsibility for handling the Refugee Legal Service which provides legal advice and assistance to asylum seekers at the various stages of the asylum process including representation before the Refugee Appeals Tribunal.

The Board produces an annual report that is based on its activity in the previous year. Statistics relating to the total number of cases, number of cases for legal advice only and number of cases for legal aid by law centre, by type of case are published as part of the report. In addition, the report also deals with administrative issues such as corporate planning, sample cases and the financial statements of the Board.

Because this scheme is restricted to that section of the population whose income is below a certain threshold, the data produced by the Board has limited value from a statistical point of view other than as serving the requirements of reporting on the activity of the Legal Aid Board. There is no potential for any further enhancement in the data produced.

Parental child abduction and recovery of maintenance payments

Ireland is party to a number of international agreements in respect of international parental child abduction and the recovery of maintenance payments from persons residing abroad. The statistics are reproduced in the DJELR Annual Report.

Neither of these sources is of major statistical importance (e.g. there were 102 cases of parental child abduction in 2001 and 34 cases of recovery of maintenance payments).

5.8 Statistical potential

Within the criminal justice system, the single greatest potential source of statistics is the Garda PULSE system. While this system is currently being utilised to generate statistics, (see the statistical Appendix dealing with crime statistics in the annual report of the Garda Commissioner to the Minister for Justice, Equality and Law Reform), it is acknowledged that the compilation of crime statistics is in need of major overhaul. The Expert Group on Crime Statistics was set up specifically to carry out this review. The output from this examination will hopefully be a set of crime statistics which is perceived by the public to be objective, timely, and compiled using best statistical practice.

While the main agencies involved in the criminal justice area (namely, the Gardaí, the Courts and the Prison Service) are autonomous, there is a high degree of interdependence between them. It is recognised within the department that this interdependence requires that their databases should be fully integrated primarily with a view to achieving operational efficiency by eliminating the duplication, gaps and inaccuracies of the current manual interfaces. A significant by-product would be improved information and statistics for policy-makers and the public.

A DJELR policy document on integration across the Justice agencies within the Criminal Justice area was agreed to in 1999. The policy, which was contained in the department's Strategy Statement 2001-2004, states *inter alia* that electronic transfers will be used where appropriate. The implementation of the policy would require, in particular, a common software platform, appropriate network infrastructure and a common data model (consistent with the PULSE system).

To date, the Justice Sector Data Network has been activated for all associated agencies other than the Gardaí while some work has been carried out on the development of a common data model. A common identifier across all systems has been identified as a key issue and the PPS Number has been singled out as the most suitable. However, an amendment to the Data Protection legislation may be required to facilitate such an approach. These issues are under consideration within DJELR at the moment.

5.9 Conclusions and recommendations

The Garda PULSE system is the key data source within the criminal justice area and is currently being examined by an Expert Group on Crime Statistics. Realising the full statistical potential of PULSE will depend to a large extent on what recommendations the group comes up with and the extent to which these recommendations will be capable of being carried out by the Gardaí.

Integration of the data sources in the criminal justice area would have a positive spin-off in terms of more comprehensive statistical output.

Realising the full statistical potential of the immigration databases may require dedicated statistical resources within that area. However, it should be possible to compile accurate stocks and flow figures for the relevant target populations on the registered non-EEA nationals database. This would be of major importance in improving the accuracy of the annual population estimates published by the CSO.

On a general level, the power to assess the statistical potential of the records of public authorities is set out as part of the functions of CSO in Section 10 of the Statistics Act, 1993 as follows:

(1) The functions of the Office shall be the collection, compilation, extraction and dissemination for statistical purposes of information relating to economic, social and general activities and conditions in the State.

(2) The Office shall have authority to co-ordinate official statistics compiled by public authorities to ensure, in particular, adherence to statistical standards and the use of appropriate classifications.

(3) The Office shall have authority to assess the statistical potential of the records maintained by public authorities and, in conjunction with them, to ensure that this potential is realised in so far as resources permit.

This power is brought into sharper focus in sub-section 31(1) of the Act as follows:

The Director General may request any public authority to consult and co-operate with him for the purpose of assessing the potential of the records of the authority as a source of statistical information and, where appropriate and practicable, developing its recording methods and systems for statistical purposes, and the public authority shall comply with any such request, in so far as resources permit.

However, the provisions of sub-section 31(1):

shall not apply to records pertaining to a Court, the Garda Síochána or the prison administration.

The Statistics Act therefore does not offer any legal underpinning to CSO specifically in relation to the administrative data holdings of agencies in the criminal justice area.

Dealing with the PULSE system in particular, the CSO is on record as stating that it is willing to provide relevant professional assistance to the Gardaí if requested to do so. This could cover areas such as:

- Technical advice on data handling;
- Adherence to statistical standards;
- Classifications;
- Data flows;
- Data quality assessment,
- Data integrity; and
- Presentation and publication of results.

For the courts and prison systems, the requirement for such professional statistical assistance may not be as pronounced as for the Gardaí.

The department is reactivating a proposal which it originally surfaced in its 1998 IT Strategy in relation to the creation of a criminal history repository. What is intended is the creation of a single location which would hold anonymised data from the information systems of the Irish

Criminal Justice System. Ideally this would require a common data model for the PULSE, CCTS and PRIS systems (i.e. shared definitions of codes). While this proposal did not enjoy much success in the past, it is likely to become more of a priority now in the context of the Department data strategy that will be required from late 2003 onwards.

Chapter 6

Department of Social and Family Affairs

6.1 Introduction

The DSFA is an important source of secondary data across a broad range of domains but in particular in relation to the provision of information on vulnerable groups. The extent to which this data can be exploited is largely dependent on the ability to link the data available within DSFA with other government departments and of course CSO statistical instruments.

Traditionally, the primary use for data held within Departments is the administration and management of the schemes and services provided. The statistical emphasis has been on the production of data that allows departments to monitor the performance of their schemes (programme indicators). However in recent years there has been a change, in particular within the DSFA with the formation of the National Office for Social Inclusion (OSI) set up as part of the National Anti-Poverty Strategy (NAPS). The OSI has been given the lead role in the development of a research and data strategy on poverty and related issues.

The primary sources of social information, especially in relation to the NAPS and NAPincl indicators, are the Quarterly National Household Survey (QNHS) and EU Statistics on Income and Living Conditions (EU-SILC). These two surveys alone provide 17 of the 18 Laeken indicators (they cover four aspects of social inclusion: financial poverty, employment, health and education). However, a significant limitation of these two surveys is their inability to provide statistical information on the main vulnerable groups in our society. Primarily this is due to the fact that these groups cannot be targeted during sample selection to ensure adequate coverage in the sample which would allow the calculation of accurate estimates from the survey. The data holdings within the DSFA have the potential to be a major source of information on these groups categorised by geography, age, family status and benefit-history. Significant value could be added to the DSFA data holdings and the CSO household survey data, if it were possible to link the sources.

In terms of data holdings (and the structure of these holdings), DSFA is perhaps in a much stronger position than many other government departments. While some of the existing systems are effectively old legacy systems (ISTS and PENLIVE), there is a clear strategy to move over to a pure relational database type structure through the Service Delivery Model (SDM). Over time, schemes will be migrated from the old legacy systems to the new SDM. This will provide the department with a unique opportunity to review the structure and content of its existing data holdings.

SDM is a ten-year programme of change. The current roll-out plan is as follows:

Phase 1	Child benefit (2002);
Phase 2	Retirement and old age contributory pension, related household benefit and bereavement grant (estimated implementation mid-2005);
Phase 3	Invalidity, widow/ers contributory pension (including occupational injury benefit widows), orphans contributory allowance and related household benefits;
Phase 4	Old age, widows and orphans non-contributory pension, carers and disability allowance, blind person's pension and related household benefits;
Phase 5	Non DSW household benefits, disablement, rent allowance, deserted wives benefit/allowance, prisoners wives allowance; and
Phase 6	Short-term schemes (sub-phases to be decided).

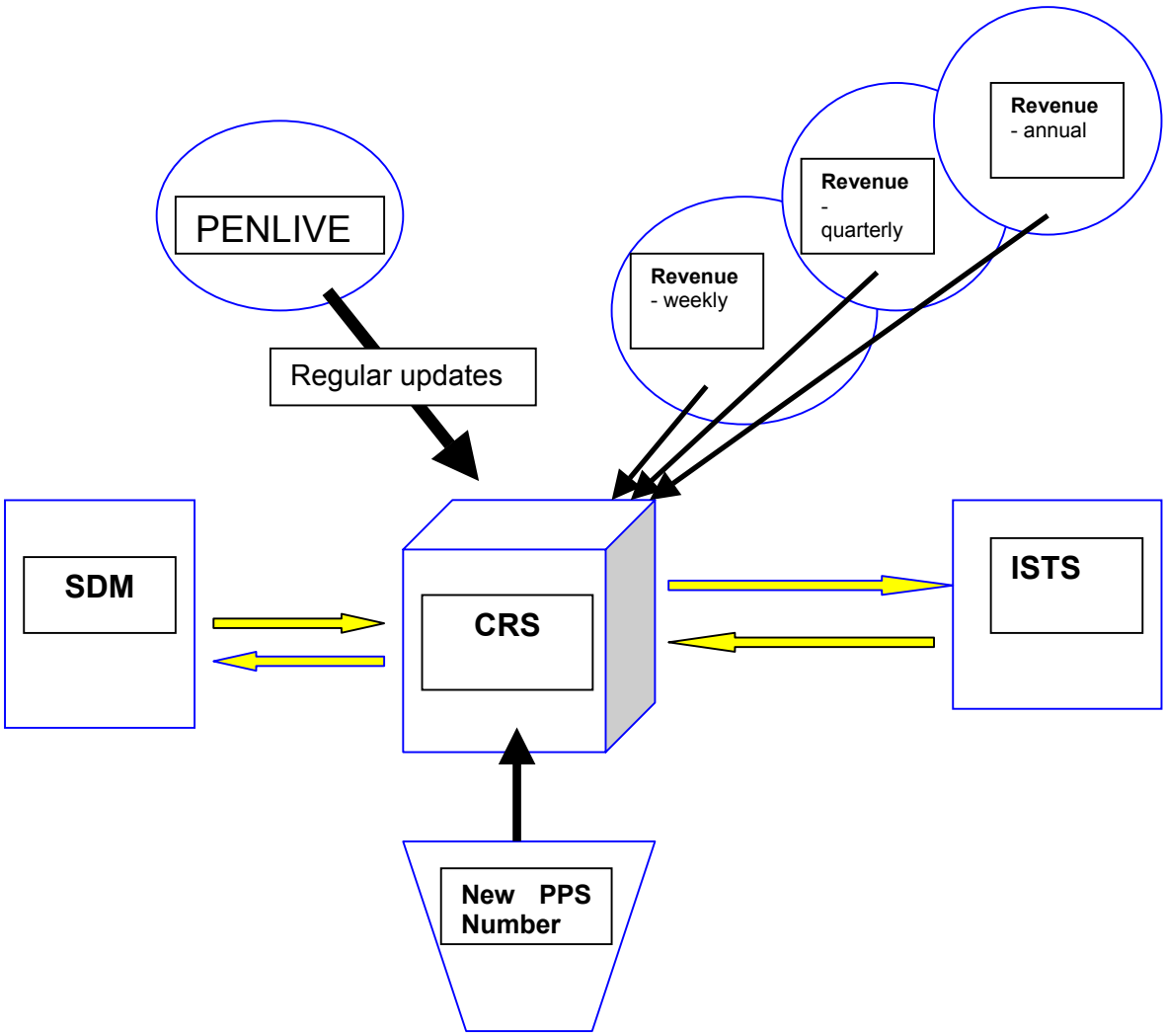
Extensive use of the PPS Number on all schemes and consequently on all systems adds considerable value to the DSFA data holdings. It is widely acknowledged that the primary key to unlocking the full potential of administrative records for statistical purposes cannot be achieved without the widespread use of the PPS Number. In terms of register development,

the DSFA are well advanced on this issue. In addition to the fact that there is now a key unique identifier (PPS Number), supplementary socio-demographic information required for statistical analysis is gathered in the vast majority of DSFA schemes. While there are a number of issues that need to be addressed in terms of standardisation etc., the key infrastructure is in place (or in the process of being put in place e.g. SDM) to allow the exploitation of DSFA data holdings for the purpose of developing social statistics.

Data sources examined

DSFA data is held on four main systems - namely the Central Records System (CRS), Integrated Short Term Schemes (ISTS), PENLIVE, and Service Delivery Model (SDM). Each system administers multiple schemes and the structure of the data holdings is consistent within any given system. For that reason our team concentrated on the systems firstly and then undertook an examination of the individual schemes within each system.

There is a substantial degree of interaction between the systems in terms the sharing/flow of data across systems. The following graphic attempts to illustrate the relevant data flows:



6.2 The Central Records System

The Central Records System (CRS) is a central repository of selected data held on different systems within the DSFA, namely ISTS, PENLIVE and SDM, and the P35 employer and social contributions data which comes from the Revenue Commissioners. CRS holds records on an individual basis and uses the Personal Public Service (PPS) Number as its identifier. The PPS Number was introduced in the Social Welfare Act 1998 to assist in the development of eGovernment and replaced the existing Revenue and Social Insurance number. The CRS is effectively a register of all PPS Numbers issued and therefore covers all persons who: pay tax; claim Social Welfare benefit or assistance; claim child benefit or receive a contributory or non-contributory pension. This effectively covers every normally resident person in the State and so is believed to have close to full population coverage. The system was developed into a relational database in 1994 and now operates in tandem with the new Service Delivery Model (SDM) and the ISTS systems in that there is a seamless data flow between all three systems. New records are added to the CRS through the issuing of new PPS Numbers by the Client Identity Services section in DSFA. The issuing of new PPS Numbers is now done following strict procedures and only on the basis of accurate and verified information. At present, all children are issued a PPS Number upon application for Child Benefit payment. Under a new system, ready to go live under the Reach initiative, this will change with the introduction of a new E-link to the GRO whereby PPS Numbers for new-born children will be assigned upon registration of birth.

CRS Data

The client table on the CRS holds a range of the key socio-demographic variables of interest for policy purposes, in particular sex, date of birth, date of marriage, marital status and nationality. Other key data of interest is held in a variety of other tables including household ID number, income data, relationship data, occupation codes, employer details and a count of jobs for persons in multiple employment. All employment and employer data is updated from files from the Revenue Commissioners. A Social Welfare Advice File is received every week, which contains details of any changes in employment status during that period for all employers and employees. Every quarter a Business Name and Address file is received containing all employer names and addresses, trading status, industry codes, cessation dates, etc. Once a year a file containing up to date data on income earned and social welfare contributions paid by, or on behalf of, all persons who were employed during the period is supplied to DSFA from Revenue. Income data on the CRS other than that covering PAYE type income is incomplete (i.e. self-employed etc.). The data on the self-employed can be accessed on the Revenue files via the PPS Number (see Chapter 7).

6.2.1 Statistical quality assessment

The quality and completeness of all the data on the CRS varies considerably depending on the data in question and the source of the data. For example, while the relationship data sourced from the SDM/ISTS systems for dependants and their families is extensive and includes all parent / child and sibling relationships, that sourced from the PENLIVE systems is less complete. The CRS is to a large extent an amalgamation of data from different systems and direct entry into the system is extremely limited. The quality of the data on the CRS is only as good as that of the data system from which it originates and this is discussed in more detail in the relevant ISTS, SDM, and PENLIVE sections.

A detailed study of the quality of the data on the CRS was outside the scope of this report. An exercise undertaken within the department in February 2002 examined the extent of missing values of some of the core demographic variables. The results are limited in their use however, as they include approximately 550,000 records subsequently identified as redundant, and a more detailed analysis would need to be undertaken before any firm

conclusion could be drawn. An extensive data cleansing exercise is now underway which aims to verify all client identity type information within each record (name against sex etc.), and to clean the existing data (remove titles from the first name field, correct overflows of data from one address field into the adjacent one, etc.). This data cleansing exercise is also aimed at addressing the issue of missing values within the client identity information. The development of the Public Service Identity (PSI) database, which will rely heavily on the CRS, will no doubt add to the business need to address additional data quality issues.

One of the more important tables within the CRS system from an analysis point of view is that which is designed to hold the detailed geographic information. However, this table is not currently populated, and as it is the only table that holds ED and street/townland codes, this means the CRS has no useful geography coding, apart from county codes. It should be noted however that the system does hold the address text strings. A similar problem exists with the client occupation table as it is not populated.

Coverage

While the CRS is believed to have almost full coverage for every normally resident person in the State, a key exception to this is elderly persons claiming state pensions that are administered at present through a claim number on the PENLIVE system. An exercise to merge these records onto the CRS was undertaken recently. As all of the various pension and long-term schemes administered by the PENLIVE system are migrated onto the new SDM over the coming years all remaining records will be assigned, or matched with, a PPS Number and merged onto the CRS. Secondly, it is possible that there are a very small number of other persons not included on the CRS i.e. elderly men or women who have never worked outside the home, never claimed a pension and never claimed any SW assistance. As the stated business aim of the CRS is to only hold information on persons necessary for administrative purposes, the Client Identity Services section of CRS administration have no plans in place to systematically add these persons as there is at present no business need to do so. However, with the development of the PSI database and the increasing use of the PPS Number by a wide range of Government bodies and agencies over the coming years, it may become necessary for all persons to be registered.

The total number of person records currently on the CRS, before any clean-up exercise, exceeds five million whereas the Census 2002 results show 3.85 million usually resident persons. The CRS does not at present systematically mark inactive or redundant records on the database.

As mentioned previously, an exercise was undertaken last year that identified approximately 550,000 inactive records (for 7 years or more). These records were marked dormant, and excluded from CRS statistics. They can be accessed and "re-activated" by all departmental staff and in fact a small number have been re-activated since being marked as dormant. The second phase of the project (still being developed) is to have all re-activations alerted to a Control Area which would follow up on these to ensure the use of the number was valid. From a register perspective, the marking of records as inactive or dormant as distinct from removing them from the database is good practice as it allows historical analysis of the data if required. This was a once-off exercise although there are now plans under the current data cleansing project to repeat the exercise and restrict the ‘inactive period’ to five years.

Initial comparisons with the results of Census 2002 (using CRS data before redundant records were purged) indicate the biggest differences are in the 23 to 40 age brackets, and in the over 70s. A more detailed comparison of numbers between the CRS file and the summary Census 2002 results of age, sex and marital status would prove useful in identifying specific groups for further investigation.

The date of death and a deceased indicator are held as part of the client information on the register, though this is only captured systematically where it is needed for administrative purposes (pensions, SW assistance stopping). It is not captured systematically for persons in employment or receiving private pensions, for example. Secondly, a date of death entry does not trigger the record to be marked redundant (as there can be 'activity' on a case even though a person is dead). Under the new E-link with the GRO, due to go live shortly, it will be feasible to have all deaths notified to DSFA and updated automatically onto the CRS, subject to the required legislation being in place. In addition, there are tentative plans in place to address the issue of historical deaths and systematically match this information onto the register, but this will be sometime in the future.

While there can be no duplicate PPS Numbers on the CRS, DSFA acknowledge that there are duplicate persons, which they divide into two types, accidental and malicious. While it is difficult to identify the latter group and their number is unknown, it is not thought to be significant. Accidental duplicates exist through the administrative practice in the past of issuing a second PPS Number if a person had lost or forgotten their first number. The data cleansing exercise referred to above aims to verify all client identity type information within each record (name against sex etc.) and, to identify duplicates and purge them. The completion timeframe for this exercise is end 2004.

6.2.2 Statistical potential

As things stand there are limitations to the use of the CRS for statistical purposes given the issues raised previously i.e. redundant records, missing PENLIVE claimants (elderly), duplicates, updates to socio-demographic variables (in particular marital status) based primarily on claims data. Many of these problems relate primarily to the historical records. Current procedures ensure full and accurate capture of the socio-demographic data for all active records which when combined with the PPS Number offers significant statistical potential.

The absence of detailed geographic data below county level has been identified as a weakness of the CRS. However, the data to populate these tables is in existence from both the Ordnance Survey, through the use of the GeoDirectory, and the actual detailed address data captured on all DSFA claim forms. The use of postcodes, which is currently under discussion, could also resolve this problem.

The CRS provides good and up-to-date income data for all employees at any given time through the P35 employer and social contributions data from Revenue. The same can not be said of the self-employed as DSFA receive a limited amount of data on the self-employed from Revenue. The provision of more comprehensive income data on the self-employed to DSFA by Revenue would add significantly to the overall income holdings on the CRS.

The DSFA data strategy, the Reach initiative and the modernisation of the GRO will drive the cleaning process of the CRS. As PENLIVE and ISTS schemes are migrated to the SDM, a detailed investigation of the data will take place and it is hoped that this will lead to an improvement in the overall quality of the data available. The ability to link with the GRO should aid the process of the identification of redundant records. The GRO link should in time also be able to provide a mechanism for updating marital status of individuals.

6.3 The Service Delivery Model

The Service Delivery Model (SDM) is a relational database system that will ultimately be used to process all schemes operated by the department. To-date, only the Child Benefit scheme (CB/SDM) has been completely migrated onto the new system. The system holds

all key demographic data such as first names, surnames, addresses, sex, date of birth, date of marriage, marital status, spouses' names, mothers birth surname, nationality and phone numbers. In addition, it holds other useful information including a school identity number for 16-19 year olds and comprehensive relationship data on all clients covering spouse, parent / child and sibling relationships. The system also holds a variable that identifies why a particular payment was stopped or extended e.g. child is in full-time education, medical certification, child has left the State, child has left household, child is in care, child has died, child has left school.

An important point to note about the SDM system, however, is that it primarily holds data necessary to process payments. While this generally covers all demographic data there are additional questions asked on all the various Social Welfare claim forms that are not captured in the SDM system. This is a common feature and a key weakness, from a statistical point of view of the Social Welfare systems generally. In the case of child benefit, the following is asked on all CB forms and not captured: previous address if moved recently, details of benefits received from other countries (country, name of payment, date of payment), employment details from work abroad (social security number, country, date last worked), immigration details if arrived from abroad (date arrived, address abroad, nationality, spouse's nationality), spouse's details (address if different, claim details, work abroad details), name and date of birth of children in hospital/home, the course, school and student identification data for children age 16-19.

6.3.1 Statistical quality assessment and statistical potential

The Child Benefit payment is unique in that there are no PRSI conditions and a claimant does not have to satisfy a means test in order to qualify. This would indicate that full coverage of all children should be expected. Current application procedures and checking systems are believed to be strict enough that duplicates are not thought to be a problem. Child Benefit is paid to a qualified person in respect of a qualified child, so essentially the child must be ordinarily resident in the State and the child must be normally resident with the person, usually the mother, for the person to qualify. Asylum Seekers, Work Permit holders or Non-EU or Non-EEA citizens must have documentation to establish their right to reside in the State, an asylum seeker card, work permit or certificate of registration from the GNIB (Garda National Immigration Bureau). Once the residency position has been confirmed these claims are dealt with in the same manner as all other claims and remain in payment as long as the child is qualified. In order to apply for a new-born child, the birth must be registered, and a birth certificate must accompany the application. At present a new PPS Number is assigned upon application for Child Benefit but the new GRO link will replace this.

The ability to identify the family unit (relationships) is particularly useful from a statistical perspective. Like many other DSFA schemes, the household unit may only be identified if 'other usual residents' are captured on the claim forms. While the capture of these additional household members may not be required for administrative purposes, their capture for social statistical purposes would be extremely beneficial. That said, it is acknowledged that the identification of the complete household may be very difficult to achieve and perhaps is not feasible in an administrative data holding environment.

6.4 Integrated Short Term Schemes (ISTS)

The ISTS system provides for the operation and administration of the DSFA short-term schemes. ISTS was the first system to be developed by the department on an integrated basis and replaced a number of systems that had been operating independently of one another.

ISTS, a legacy system, has been in operation since 1994 and holds information on claimants of Unemployment Benefit and Assistance, Disability Benefit (DB), Supplementary Welfare Allowance (SWA), Maternity Payments, Occupational Injury Benefit (OIB), Pre-Retirement Allowance (PRETA) and payments under the Back to Work, Enterprise and Education schemes.

There are three scheme types within ISTS. ISTS SWA relates to SWAs which cater for persons ineligible for other DSFA schemes (or pending these payments) or those in need of emergency payments and is administered by Community Welfare Offices. ISTS Unemployment covers Back to Work, Enterprise and Education, UA, UB and PRETA (CSO receives aggregated data for the number of claimants for these schemes at Local Social Welfare Office level for use in compilation of the Live Register). ISTS Disability covers schemes operated for cases of temporary illness or disability, including those caused as a result of accidents at work and maternity benefit. The ISTS files are the same for all three scheme types however certain variables have access restrictions according to type of scheme. A different front end faces operators of the three different scheme types. An important distinction within the system is between open and closed claims.

The PPS Number is the unique identifier on the ISTS file and allows relevant Client Identification data updates to CRS (name and address, date of birth, marital status). The socio-demographic client data on ISTS is of a high quality given the links to CRS via the PPS Number.

As is the case with PENLIVE, in assessing the ISTS data holdings it is important to draw a distinction between what is collected (i.e. on the claim forms) and what is captured (i.e. what gets input to the ISTS system from the claim form).

6.4.1 Statistical quality assessment

Good quality socio-demographic client data are available on the ISTS system. There is a considerable level of useful socio-demographic information captured on the system which is not captured on PENLIVE (nationality, breakdown of means, incapacity codes, claim category and occupation codes).

Given the short-term nature of the schemes, in general ISTS gives poor coverage of the various population groups. There are no reference populations for comparison with ISTS SWA and ISTS Disability (excluding maternity benefit).

6.4.2 Statistical potential

The incomplete transfer of useful data from paper to system is common throughout the ISTS schemes and represents a significant loss of information to the department that may not only be useful within the department but to society as a whole. This is not a problem in relation to the clients' personal details as these should be available on the CRS. However, there are other relevant variables that should be examined. For example, all educational attainments (second and third level courses) are collected on the claim form for the Back to Education Programme. However, this information is not captured by ISTS. Also, ISTS SWA schemes collect data specific to the nature of the scheme which is not captured but would be of interest (e.g. weekly costs of childcare and registration details of childcare facilities are collected for the Creche Supplement scheme). In assessing the usefulness of such data it should be stressed that the coverage of various schemes may only relate to restricted groups.

ISTS and its links to CB/SDM through CRS enables the identification of the family unit and allows the derivation of relationships between family members. Like CB/SDM the

identification of the household is not possible. Additionally, it is important to note that the system caters for the capture of data on non-dependants (which would aid the identification of the household) but there is no guarantee that this data is collected on all claims.

6.5 PENLIVE

PENLIVE is used to manage schemes relating to pensions, lone parents, free travel, household benefit packages, invalidity, certain allowances (Carers, Disability) and other long-term claims paid via a book system. It is an old legacy system with data dispersed over 11 to 12 separate files. However, these files can be linked using one of three unique identifiers used in the system. These identifiers are PPS No, client/pension number and claim number. The PPS Number is only available on claims post 1994 and as the system was developed in the 1980s, the PPS Number is not the main identifier on the system. However, almost 95% of all 'live/relevant' claims now have the PPS Number attached to the relevant record.

Administrative data are normally collected with the operation of the organisation in mind rather than with a view towards statistical analysis. This is especially true of the PENLIVE system which is effectively a payments driven system. In assessing the PENLIVE holdings it is important to draw a distinction between what is collected (i.e. on the paper claim forms) and what is captured (i.e. what gets input to the PENLIVE system from the claim form). In general, the only data captured is that which is required to drive a payment.

6.5.1 Statistical quality assessment

The socio-demographic client data on PENLIVE is of a high quality. In particular, details of name and address, date of birth, marital status/date of marriage and date of death are accurate. These fields are updated to CRS for confirmed PPS Numbers. PENLIVE holds a long time series of data, as records are not systematically erased from the system.

There is a limited amount of client data held on the system with details of dependants only added in the 1990s. Compared to ISTS/SDM, there is a considerable level of useful information not captured on the system (nationality, breakdown of means, incapacity codes, position prior/claim category and occupation codes).

While it is possible to identify claimants with a disability via the scheme code, there is little or no information on the disability. Furthermore, in relation to claimants with a disability, and specifically those claiming through the Disability Allowance, the fact that this scheme is means tested does have an impact on the possible usefulness of the data due to the associated coverage issue.

A fundamental aspect that needs to be addressed within PENLIVE is the inability to identify the date a claim stopped (not recorded), which makes searches and register matches difficult. Client data for stopped claims may be of low value, depending on the date the claim stopped. The group with no PPS Number is getting smaller so this is less of an issue with the passing years.

6.5.2 Statistical Potential

The incomplete transfer of data from paper to system is common throughout the PENLIVE schemes and represents a significant loss of information that may not only be useful within the department but to society as a whole. This is not a problem in relation to the clients personal details as these should be available on the CRS. However, there are other relevant variables that should be examined. These include residence in the state, non-dependent spouse and child details. Significant value-added could accrue were this data to be captured

especially given the existence of the PPS Number and the possible links to data sources such as EU-SILC.

The ability to identify the family unit is a key analysis variable in terms of social statistics analysis. Currently it is not possible to identify family units directly within PENLIVE as only data on dependants (whether adults or children) is actually captured on the system and there are no relationship codes. The collection of data relating to the spouse/partner and children, regardless of whether deemed dependants or not, should be given serious consideration. In particular, consideration should be given to the collection and capture of all the socio-demographic data plus the PPS Number of the spouse/partner and children.

6.6 Review of DSFA claim forms

During the first contact with prospective customers a detailed claim/application form is completed. This varies by scheme, covering a broad range of issues including client personal details, employment related information and information relevant to the calculation of an individuals means (where required).

DSFA claim forms have a similar look and feel. There are variables that appear on the majority of the claim forms, however, these are sometimes asked in different ways or sometimes a related but different question is asked. For example, on the Disability Allowance claim form the applicant is asked for his/her “Nationality”. However, on the Blind Person’s claim form, the applicant is asked - “What country were you born in?”

From a statistical perspective one could argue that a standard set of questions, regardless of the scheme, should be considered for the client and spouse, where appropriate. The collection of demographic details for the remaining household members, including relationship codes, would also allow the identification of the family unit and household. However, it has already been acknowledged that the achievement of the latter may not be possible in an administrative data holding environment.

6.7 Classifications and coding practices

The classifications for any given variable are not standard/common across the systems, however they are within the systems (SDM/ISTS/PENLIVE). A notable exception relates to the One Parent Family Payment claim form where an additional category is used to identify a claimant whose spouse/partner is a prisoner. While the need to identify this group is clear, it is important to retain the distinction between classifications and questions. Effectively the marital status of a claimant in this situation is “married”. A subsequent clarification of this marital status should be addressed through an additional question and not through a change in the classification.

A decision to adopt a standard set of classifications to be used by all data collecting agencies (public and private) should be taken as this is a key issue in relation to the successful analysis of data (micro and macro). The classifications to be used for the main socio-demographic variables should be those used by the CSO.

ISTS codes sex as M-male and W-female while SDM codes sex as M-male, F-female, and U-undetermined. PENLIVE uses just two categories M-male and F-female. The county details within addresses are not coded in ISTS/SDM but during the up-load procedure to CRS a code is assigned. In general, occupation coding by the DSFA does not follow any current internationally recognised classification (e.g. Standard Occupational Classification).

The classifications and coding practices applied to the data held within the CRS are in effect inherited from the source data holding (ISTS, PENLIVE, SDM, Revenue, etc.). As such, the standardisation in classifications and coding practices across all systems both within DSFA and outside, will eventually be reflected within the CRS.

As the schemes in ISTS/PENLIVE move over to SDM, the opportunity to standardise the classifications should be taken. The classifications to be used for the main socio-demographic variables should be those employed by the CSO.

Within ISTS Disability (which covers Disability Benefit, OIB, Maternity Benefit) there is a comprehensive incapacity classification used which provides details on the type of illness/disability being endured by the claimant. If this was extended to all illness/disability related schemes, it could add significant value to the DSFA data holdings on these related topics.

6.8 Cross cutting aspects

6.8.1 CRS

The CRS has the potential to act as a hub allowing the merging and matching of data (administrative and statistical) from different sources through the use of the PPS Number, subject to data protection legislation.

Legislation enacted in 2002 provides for the development of a PSI database as part of the Public Services Broker, under the Reach Initiative. It will contain a limited set of key demographic variables and is being developed to assist government agencies in the authentication of identity as well as providing some additional services e.g. find a PPS Number, find a customer for a PPS Number etc. The PSI data consists of: surname, forename, date of birth, place of birth, sex, mother's birth surname, birth surname, address, nationality, date of death, and in the case of a person who is under 18 years, mother's PSI and father's PSI. The initial strategy is to provide a copy of this data from the CRS, which holds all the required data with the exception of place of birth.

Part of the role of the Client Identity Services of CRS is to assist other organisations in the use of the PPS Number. Many agencies have adopted the PPS Number, mainly by directly requesting the PPS Number from their customers. Projects were undertaken to supply PPS Numbers to the Education post primary database and the General Medical Services (Payments) Board in respect of medical card holders.

In addition, the PPS Number is now used by the following bodies: Central Statistics Office, Revenue, the Department of Education and Science, the Department of the Environment, Heritage and Local Government, some Local Authorities, the Irish Cervical Screening and Breast-Check Programmes, the Government Special Savings Incentive Scheme, the Department of Health Immunisation Programmes, the General Register Office, the General Medical Services (Payments) Board, the Medical Card and Drugs Payment Schemes and the Department of Agriculture and Food.

6.8.2 SDM

Maintaining accurate data on an individual from birth to death has significant statistical opportunities in terms of capturing socio-demographic information in conjunction with a unique identifier. The CB/SDM provides the information based on the first point of contact with an individual and starts that process of information gathering at the micro level. The use of the PPS Number as the identifier allows analysis of data from a myriad of administrative

and statistical sources. Examples of the type of analysis that could be achieved include the link between education and earnings, indicators of early mortality (education, income, etc.), impact of family situation on life events etc.

6.8.3 ISTS

ISTS holds data on some of the key social groups where there is currently a considerable lack of information e.g. low-income families, homeless, asylum seekers. The claim category code used for Supplementary Welfare Allowances enable the identification of these groups although it is important to keep in mind the qualification rules for any given scheme. Within ISTS, it is possible to undertake analysis of these groups given the broad range of socio-demographic data available. The income related data collected for means purposes can also provide some interesting analysis.

The DJELR have indicated a need for Unemployment Assistance and Benefits data to be broken down by nationals and non-nationals. Nationality of applicants for ISTS Unemployment and ISTS SWA schemes are collected but not captured in all cases. This is a problem but the data required to meet this demand does reside within the department.

6.8.4 PENLIVE

PENLIVE holds data on some of the key social groups where there is currently a considerable lack of information e.g. one parent families, carers, the disabled and the elderly. The scheme codes used within PENLIVE allow for the identification of these groups although it is important to keep in mind the qualification rules for any given scheme. Within PENLIVE itself, it is possible to undertake analysis of these groups given the broad range of socio-demographic data available, and the income related data collected for means purposes can also provide some interesting analysis.

The EU has agreed a set of 18 common indicators (Laeken indicators) covering four aspects of social inclusion: financial poverty, employment, health and education. The QNHS and EU-SILC combined provide the source data for 17 of these 18 indicators. The main challenge, facing all involved in the provision of this data is how it can be disaggregated to provide these indicators at the level of the different marginalised/vulnerable sub-groups. It is conceivable that the combination of administrative data and the CSO household surveys could provide us with an opportunity to derive these indicators for these sub-groups.

The QNHS, while concentrating primarily on labour market issues, also covers education related issues in a comprehensive manner. Changes to the survey in 2002 and further changes planned for the latter part of 2003 will provide a wealth of information on this topic. Again combining this data with the PENLIVE data holdings could provide some useful insights into the groups mentioned above.

6.8.5 General

There are significant value added opportunities to be exploited by linking PENLIVE/ISTS holdings with the QNHS and/or EU-SILC via the PPS Number. The labour market data collected through the QNHS (note the PPS Number is not yet collected in the QNHS) combined with the poverty/deprivation data collected through EU-SILC, could allow a significant level of analysis at the micro level and subsequently the macro/aggregate level (obviously dependent on achieving an adequate sample).

One of the advantages that an accurate and up to date CRS file would provide is the ability for the CSO to review how samples are stratified and selected for the main household

surveys. This could allow for more targeted sampling of specific groups and thus the provision of valuable information that heretofore has not been available from these sources.

The ability to identify the disabled from DSFA data has been raised by a number of different organisations. This study indicates that the disabled population would be difficult to identify based on DSFA data holdings alone. In effect, the disabled are covered in a number of different DSFA schemes including Occupational Injury Benefit, Disability Benefit, (both short-term), Disablement Benefit, Disability Allowance and Invalidity Pension. A disabled person can move between these schemes depending on the duration of the disability, the employment status of the individual and their contribution records. Add to that the fact that a number of the schemes are means tested, it is difficult to assess the coverage of the disabled population within DSFA data holdings.

There are plans to investigate the feasibility of identifying those with disabilities within the QNHS. The need for the identification of the disabled is well documented and if the inclusion of a categorical question on this topic is successful within the QNHS then data on the ILO status, PES status, education levels, etc. of individuals identifying that they have a disability will be possible. Using the QNHS as a benchmark figure could add significant value to administrative data holdings on the disabled.

The Health and Safety Authority have expressed a strong interest in obtaining data on accidents in the work place. Injury benefit, one the benefits available under the Occupational Injuries Benefits scheme covers short-term (up to 26 weeks) payments made to persons unable to work due to an accident in the workplace or illness contracted due to the type of work the person performs. Data on the numbers of claimants and their disability, at a point in time and flows on and off the scheme, should be readily available from the ISTS file. The nature of the disability/incapacity is collected and coded in the system.

It should be noted that Injury Benefit is a short-term scheme. If the disability persists the person would become entitled to Disablement Benefit which is not recorded under ISTS (Disablement Benefit data is held on a separate system but the type of data held is of little or no use from a statistical perspective). Also, the schemes are restricted to certain insurance classes and also to a prescribed list of occupational illnesses.

6.9 Overall conclusions and recommendations

The DSFA are well advanced in the development of a technical infrastructure that will enable the exploitation of their existing data holdings. The task now facing the Department (in common with other departments) is to develop their data holdings in such a way that will allow meaningful analysis to better inform social policy.

The emphasis of the study has been to highlight the statistical potential of existing administrative data holdings with a view to the development of social statistics. This can only be achieved through the full exploitation of all government data holdings and in many cases the linking of multiple data sources spread across numerous government departments.

During this study, it was difficult to get an in-depth statistical analysis of the data holdings. Realistically, the achievement of this would require much more time and access to the data holdings themselves. In certain circumstances the presence of a variable does not automatically infer that the variable is useful from a statistical perspective as one would need to be able to analyse the quality of the coding applied to the variable e.g. nationality. This study did not allow for analysis of this type.

In terms of the DSFA, we have identified a number of conclusions and recommendations that will contribute to the achievement of the overall strategy:

(i) Priority should be given to the cleaning of the CRS;

(ii) Geographic coding of data on the basis of the Ordnance Survey GeoDirectory should be considered;

(iii) The identification of the household unit and the relationships within the household are important analysis variables from a social policy perspective. In reality it is very difficult to see how these variables could form part of administrative data sets in the short to medium term. Even as a long-term objective it may be considered ambitious;

(iv) A standardised core set of socio-demographic questions (covering details such as address information, sex, date of birth, marital status and nationality) should be asked on the claim forms for all schemes the first time the PPS Number is captured on the system. Thereafter, those details subject to change, e.g. marital status, need only be updated;

(v) An additional set of variables that would be of considerable interest from a social perspective, should be considered for inclusion on all claim forms (where appropriate). These include usual place (country) of residence, occupation, Principal Economic Status (PES), living alone indicator, and a standard set of means related questions;

(vi) The identification of vulnerable groups is a corner-stone in terms of the development of social statistics. To meet this objective, consideration should be given to the inclusion of categorical type questions that would identify ethnic groups and those with a disability;

(vii) Specifically in relation to disability, consideration should be given to the extended use of the incapacity classification (where feasible and appropriate) currently used within ISTS to all schemes capturing any illness/disability type data;

(viii) Coverage is a difficult issue to address given that many schemes are means tested. This effectively creates an administrative population as distinct from an actual population;

(ix) Use of agreed standard classifications should be applied across all systems and schemes; and

(x) Electronic capture of all data that is collected on the paper forms.

Chapter 7

Office of the Revenue Commissioners

7.1 Introduction

The Office of the Revenue Commissioners (Revenue) administers the State's tax system. Its activities mean that the Office interacts with almost every citizen and business in the State. As a result, the administrative data it collects is a very extensive and useful source of statistical information.

The SGSES report identified several data items within Revenue's administrative systems that might provide useful additional social statistics. The remit of the SPAR team was to examine these in more detail and to assess their statistical potential. While the primary focus of the exercise was on social statistics, the SPAR team was also asked, if possible, to look at other data holdings to assess their usefulness for statistics. In the event, the limited time available meant that this was not possible and this study is largely confined to those data sources identified in the SGSES report as being of social interest. However, Revenue's data holdings primarily contain financial and business information and it is expected that their primary usefulness will be in the area of economic statistics. It is intended that this aspect will be examined later in more detail in an additional study, which will examine the potential of administrative records for economic and environmental statistics. This report therefore only covers this issue in a cursory manner.

A total of 22 separate sources were identified in the SGSES report. The majority of these were derived from the income tax system which covers both PAYE and self-employed. After meeting with officials from Revenue, the SPAR team agreed that for the social statistics study only two main data systems required in-depth analysis. These were the income tax and the capital tax systems. The information collected and the conclusions of the SPAR team are reported in this chapter. For completeness, the other data holdings mentioned in the SGSES report are also described.

7.2 Income Tax system

The Income Tax system comprises individuals who are in receipt of income in their capacity as employees, as self-employed or as investors. The method of collecting and assessing taxes payable depends on the employment status of the taxable person.

In the case of employees, income tax is assessed and collected through a combination of PAYE and on-line review (OLR). The operation of the PAYE system is enabled primarily through the compliance of employers. An on-line review is undertaken on a more ad hoc basis and may be carried out by Revenue on its own initiative or at the request of the taxpayer.

Income tax of self-employed persons and investors is generally collected on an annual basis through a self-assessment system. This is a compulsory system, and stiff penalties and interest payments are levied for non-compliance.

Up until 2001, the income tax year differed from the calendar year, but the two have been aligned since January 2002. However, self-employed persons are still allowed to choose an accounting year that differs from the calendar year. This limits comparisons between incomes of self-employed persons and those on PAYE.

A more detailed description of the income tax data sources follows.

7.2.1 PAYE file

Under the PAYE system, employees have income tax deducted directly from their wages. The employer, on the basis of instructions issued by Revenue, calculates the income tax payable by each employee. The amounts vary and are dependent on entitlements to allowances, tax credits, tax rates and tax bands. These are generally determined by an individual's marital status.

The employer is also obliged to deduct a Pay Related Social Insurance (PRSI) contribution from the employee's gross wages. The employer also pays an additional top-up PRSI contribution in respect of each employee. The rules for calculating PRSI are complex and the contribution depends on a number of variables including the type of employment and the level of wages. Most employees are covered by Class A1 which entitles them to access a number of schemes under the Social Welfare Acts, including unemployment, disability and maternity benefits and old age and widow/widower pensions. These schemes were covered in Chapter 6. In administering contributions under these schemes, employers and Revenue act as agents for the DSFA.

An employer remits PAYE and PRSI deductions to the tax office every month. The remittance is supported by a statement (Form P30), which indicates the total amount of PAYE and PRSI deducted in the month. This information is provided only at an aggregated level and it is not possible to determine the amount paid per individual employee.

However, an employer is also obliged to submit a Form P35 within six weeks of the end of the tax year together with any balancing PAYE or PRSI payment that may be due in order to balance with the twelve monthly P30 forms. Form P35 provides detailed information on each individual employee, including:

- PPS Number;
- Total pay earned in respect of the employment;
- Total PAYE deducted (or refunded);
- Employee's share of PRSI;
- Total PRSI deducted (includes employer's and employee's share);
- PRSI class and number of weeks in insurable employment; and
- Where an individual commences or leaves the employment during the particular Income Tax year, these dates are also provided.

If an employee's PPS Number is not known, the employer must provide details of the employee's date of birth and mother's maiden name in order that a number can be supplied to the employee subsequently. This number is generated by Client Identity Services section of the DSFA (see Chapter 6).

The form P35 is an important document, as it is needed to determine an individual's tax credits and social welfare entitlements. It is understood that there is excellent coverage, with defaulting employers (less than 4%) being pursued rigorously.

An employer is also obliged to issue a form P45 when an employee ceases work. This form is in multiple parts and contains similar information to that contained on the P35. The P45 is used by Revenue to update the PAYE file with the employment cessation date of the particular employee. If the employee gets another job, the same form P45 is used to record the commencement date of the new job. This information is primarily used for tax and social welfare administration purposes, but a by-product of the P45 system is the ability to generate statistics of the numbers of persons entering or leaving an employment. These inflows and outflows have been analysed quarterly for a number of years. When the Labour Force Survey was undertaken annually, they were used as short-term indications of trends in employment between survey dates. The Quarterly National Household Survey now provides such information.

7.2.2 On-Line Review file

In addition to the P35 return made by employers, taxpayers can also complete and submit an annual tax return (Form 12), which includes details of all sources of income and of the claims for allowances or deductions to which they may be entitled. Nowadays, some allowances are deducted at source while many others can be obtained without completing a full tax return. The allowances and credits that taxpayers can claim include:

- A blind person's credit, including a credit for having a guide dog;
- Age allowance, once taxpayer reaches 65 years of age;
- Incapacitated child credit, where the taxpayer looks after a dependent child;
- Dependent relative credit, where the taxpayer looks after an elderly or infirm person or a widowed parent;
- Home carer's credit to a spouse who works at home to care for children, an aged person or incapacitated person;
- An allowance for employing a person to take care of an incapacitated individual;
- Medical expenses incurred over a specified threshold;
- Medical Insurance premiums (relief is granted at source from 6 April 2001);
- Mortgage interest relief (relief is granted at source from 1 January 2002);
- Premiums paid on long-term care insurance policies which have been taken out to provide for future care;
- Needs of certain individuals (relief is granted at source);
- Rents paid by tenants; Persons aged over 55 years can claim double the relief allowed to under 55 year olds; and
- Fees paid to third level colleges and for approved training courses.

Revenue has for many years moved away from requiring each PAYE taxpayer to submit an annual tax return. Such returns are now principally submitted by taxpayers, who are seeking reviews or who are specifically requested to do so by the tax inspector. Revenue issues each year approximately 100,000 to 120,000 Tax Return Forms (Forms 12) to selected

PAYE cases. These are targeted on the basis of their income levels, income sources or the reliefs they claim. Claims for tax credits and tax reliefs, etc. can also be processed over the telephone, in public offices or by correspondence (including email), and no longer need to be supported with a tax return form. However, the PAYE data sets are automatically updated with all information obtained in such contacts.

The assessment and recording of incomes and taxes in the case of single individuals is straightforward. The arrangements in the case of married persons are more complex. Under the income tax system, married couples can choose to be jointly assessed, separately assessed or singly assessed. In the case of joint assessment, only one return of income is required which contains details of the respective incomes of both spouses together with any claims for allowances and credit

7.2.3 Direct Assessing (Self-assessment) file

The Taxes Acts require that individuals who are in receipt of self employed incomes or investment incomes are obliged to submit a completed return of income (Form 11) each year, from which their liability to both income tax and PRSI is assessed. This is known as the Self-Assessment system. Directors owning more than 15% of equity in a company (proprietary directors) are also covered by this system. There are strict deadlines and reporting requirements and failure to comply carries interest and surcharge penalties.

Self-assessment was introduced in the Tax year 1988/89 and has therefore been operating for almost 15 years. The taxpayer makes an interim payment of income tax, referred to as Preliminary Tax, followed by a payment of any outstanding liability after the submission of the income tax return.

The tax return provided by the self-employed is more detailed than that completed by employees and provides for a further breakdown of the types of income. The returns will also have copies of accounts for the business annexed to them. However, details of the claims for personal allowances and tax credits are identical to those on the PAYE form.

7.2.4 All Persons file

Revenue has a Forecasting and Statistics Branch that analyses the income tax returns and publishes results in the annual Statistical Report of the Revenue Commissioners. To facilitate their analyses, the Branch combines the data collected on the P35, the On-Line Review and the direct assessing systems to produce an 'All Persons file'. Duplicates are removed from the P35 file where similar details have been separately captured in the on-line review and directly assessing data sets. The 'All Persons file' captures virtually all the basic information that is collected as part of the income tax system. The SPAR team therefore decided that it was this combined file that would be examined and assessed for quality.

7.2.5 Statistical quality assessment

Scope of income variables

The scope of the income variables on the income tax system is restricted to those components covered by the Taxes Acts. It is therefore not an exhaustive list of income from all sources. In particular, gross income as defined for tax purposes does not include, or only partially includes, the following items:

- Social welfare assistance payments, maternity and child benefits and payments from non-contributory schemes - gross income does, however, include social

welfare payments made under contributory schemes such as unemployment benefit and contributory pension schemes;

- Interest received which has been subject to deposit interest retention tax (DIRT) or interest which has been paid by An Post from their Instalment Savings Schemes;

- Tax exempt incomes such as stallion fees, earnings of writers, composers and artists and profits from commercial forestry;

- Statutory redundancy payments and certain lump sum payments made upon cessation of employment; and

- Employees' superannuation contributions - income reported for tax is net of such contributions.

Population coverage

In terms of the numbers of taxable cases, the PAYE data set is virtually complete and has no duplication. There is a high level of compliance with about 96% coverage. This data set is updated annually using information from the P35 returns made by employers. Almost 50% of these returns are made within three months of the end of the income tax year. The balance is submitted within six months and outstanding returns are pursued rigorously. Data are archived for many years.

The coverage of the self-employed direct assessing system is more uncertain. It is believed that the detailed statistics are based on about 90% of all self-assessable taxpayers.

There is low coverage of incomes earned by persons from non-PAYE sources who are not systematically or annually processed in the tax system because they have consistently low individual incomes. Such cases are not routinely assessed to income tax and the income updates are captured only when the inspector of taxes intermittently requests the completion of a self-assessment tax return. If that return is judged appropriate to go through the formal tax assessment process, even if the ultimate outcome is a zero tax bill, the income particulars will be captured on computer record. While required under the Taxes Acts, these forms are not formally pursued unless the inspector wishes to investigate a particular taxpayer.

Completeness of records

All the information reported by employers and employees is recorded comprehensively and accurately on the data files. However, some information in the income tax system may be dated or incomplete. As explained earlier, some PAYE taxpayers do not submit an annual income tax return and the personal details of these taxpayers and information on allowances and claims may not always be up to date.

In the PAYE system, some non-PAYE incomes may not be recorded gross. When PAYE taxpayers have non-PAYE incomes, Revenue sometimes collects the taxes due on these incomes by reducing tax credits and increasing the PAYE income tax contributions. These additional sources of income are therefore not captured gross in the system unless the taxpayer also submits an income tax return that is processed by the tax inspector using the On-Line Review system. It is not possible to quantify the amounts of income thus excluded but it is unlikely to be significant because the additional incomes need to be low in order to be collected through the PAYE system.

Certain key social variables, such as age or date of birth, are not included on the tax files. These are of no direct interest for tax assessment and are therefore not captured or recorded in the tax system. There is also a difficulty with the identification of gender, in the case of jointly assessed married couples. When both spouses earn, either spouse may complete the income tax form and include details relating to 'Self' and 'Spouse' rather than 'husband' and 'wife'. This means that it is not possible to identify from the tax statistics files, which of the spouses' incomes belong to the wife and the husband.

Income details can also relate to different time periods. As mentioned earlier, self-employed persons are allowed to report their trading profit for an accounting year, which often differs from the calendar year. This, therefore, provides limitations when making comparisons between incomes earned by self-employed persons and employees, whose incomes are reported on a calendar year basis.

The Inflow and Outflow statistics mentioned above, and referred to in the SGSES report, have also sometimes been found to be weak indicators of short-term movements in and out of the work force. There are a number of reasons for this, including the fact that the correct date of commencement or cessation may not be entered onto the system until the annual form P35 has been processed. This can be up to six months after the end of the income tax year.

7.3 Particulars Delivered (PD) Form – ST 21

A completed form ST 21 (Particulars Delivered) must be submitted to a Stamp Duty Office in respect of:

- All documents giving effect to the transfer of any building and land or any interest in such property;
- The grant of any lease for a term exceeding 30 years; and
- The assignment of a lease where the remaining term exceeds 30 years.

A Particulars Delivered stamp must be impressed on the document. This requirement applies even where no stamp duty is payable. There are, however, two exceptions:

- Any transaction creating a joint tenancy between spouses to which Section 14 of the Family Home Protection Act, 1976 relates; and
- A conveyance, transfer or lease by or to a housing authority to which Section 8 of the Housing (Miscellaneous Provisions) Act, 1992 relates.

Form ST 21 collects information on the property transferred or leased and of the parties to the transaction. The details provided include name, address and tax reference number of the transferor (or lessor) and of the transferee (or lessee). The PPS Number is required for individuals and the Corporation Tax number is used for companies. The form also provides the address (including townland for rural cases) of the property and the folio number, where available. The type of property transferred is broken down into the following categories:

- Private house/apartment (with a separate tick box to identify new properties);

- Agricultural land (separate tick box to indicate if dwelling is included);
- Non-agricultural land;
- Commercial/industrial premises; and
- Other property.

For the last two categories, a further brief description of the type of property must be provided in writing.

If the property being transferred exceeds one acre, the land area must also be specified. In the case of a sale or the assignment of a lease, the full amount paid is reported on the form. In the case of a lease, information is provided on both the annual rent and the premium paid, if any. There is also a separate identifier to indicate if the property has been transferred at less than open market value.

Statistical quality assessment

This is a good comprehensive source of information on transfers of real property because the documents need to be stamped for legal purposes. Almost 130,000 returns were submitted last year. Data from 2001 to January 2003 have been electronically captured and the transfer of these data into the Ingres-based Stamp Duty Administration System (SDAS) is in the process of being tested. All information provided on the ST21 form is captured either by SDAS when a deed of transaction is presented for stamping or by separate data input of the ST 21 forms. The linkup between the ST 21 data and SDAS is facilitated by a common identification number originally generated in SDAS. There is no coding of variables or standardisation of formats of names and addresses.

7.4 Gift/Inheritance Tax Self-Assessment return (IT38)

A beneficiary of a gift or inheritance is obliged to submit a return within four months of the valuation date where the aggregate value of all taxable gifts and inheritances exceeds 80% of a group threshold. For the purpose of Capital Acquisitions Tax (CAT), otherwise known as Gift and Inheritance Tax, the relationship between the person who provided the gift or inheritance (the disponer) and the person who received the gift or inheritance (the beneficiary), determines the group threshold. The three group thresholds operating in 2003 were as follows:

Table 7.1 Gift/Inheritance tax thresholds

Group	Relationship to the disponer	Threshold in 2003
I	Son or daughter or child of deceased son or daughter. Also includes parents who take an inheritance from a deceased son or daughter.	€441,198
II	Lineal descendant, ancestor or brother, sister, niece, nephew or child of a brother or sister of the disponer.	€44,120
III	Anyone not qualifying for group I or II.	€22,060

Form IT38 collects information on both the disponer and the beneficiary. Details collected include the PPS Number of both parties, name and address, country of domicile, relationship to the disponer, the value of the property being disposed with business assets and agricultural assets identified separately.

Beneficiaries can claim exemptions and reliefs. For example, the first €3,000 taken as a gift in any one year from a disponer is exempt. Spouses are not liable to CAT regardless of the level of the benefit. There is also no liability to tax when inheriting a dwelling house where it has been the main or only residence of the recipient for a period of three years and remains so for a further six years. Other exemptions and reliefs apply when a beneficiary inherits agricultural property and certain business assets.

Statistical quality assessment

The information captured is of good quality and is complete in respect of its target population. However, the high threshold and various reliefs and exemptions applying mean that coverage is limited.

7.5 Inland Revenue Affidavit form CA24 and CA4

Inland Revenue Affidavit form CA24

The form CA24 is completed by the executor of a will in order that grant of Probate can be taken out. The procedure is to complete the affidavit providing a schedule of assets of the estate, excluding assets passing on joint survivorship (e.g. joint ownership by spouses), and listing the names of the beneficiaries under the will. The affidavit is sworn in at the grant of probate office. The form is then certified and passed to Capital Taxes branch, which assesses any inheritance tax liability on the assets of the estate. If a liability arises, the beneficiaries are obliged to submit their calculations and payment of inheritance taxes within four months.

The only data that is currently being captured electronically in capital taxes branch is the gross and net value of the estate's assets.

Form CA4

This form is completed by anyone entitled to benefit from a financial account held in either joint names with a deceased person or in the sole name of the deceased person, where it is agreed that either an original or amended form CA24 affidavit is not required. Once Revenue is happy that no liability to inheritance tax arises, a certificate is issued to enable the financial institution to release the funds to the beneficiary.

Statistical quality assessment

Form CA24 captures only a portion of the value of assets passing on death because it does not cover assets passing on joint survivorship.

Data from the form CA4 is not recorded electronically. It is an administrative document required for clearance purposes and applies in only a limited number of cases.

7.6 Data sources for economic statistics

The primary focus of the SPAR project was to analyse the usefulness for social statistics of the administrative data systems listed in the SGSES report. However, the terms of reference also requested SPAR teams visiting Departments to investigate, as far as possible, other data sets that could also be a useful source of statistics.

It has long been realised that Revenue's data holdings are an invaluable source of information for the development of economic statistics. Indeed, there is a long history of co-operation between Revenue and the CSO in this area. Ireland was one of the first countries in the world to compile National Accounts, based on information collected through the tax system. Much of this information, including data from the Corporation and Income Tax systems, continues to be used today. Another traditional source of economic statistics is the import and export customs declarations, which are used to compile the State's official external trade statistics. Since the introduction of the Single Market and the simplification of customs procedures for movements of goods between EU Member States, the VIMA branch of Revenue in Dundalk conducts the Intrastat survey for intra-EU trade and provides the data to the CSO for publication.

Nonetheless, it is recognised that other Revenue data holdings could also be better utilised to produce economic statistics and possibly reduce the existing reporting burden on enterprises. The CSO has never used the PAYE system as a source of information on earnings for the National Accounts but is now planning to investigate this possibility. The SPAR team is also aware that the CSO and the Revenue Commissioners are currently discussing the possibility of using the PAYE system to update employment records on the CSO's business register on a quarterly basis. This will significantly improve the quality and usefulness of the business register. It is intended that this new system will be implemented before the end of this year.

Another source whose potential usefulness is not being fully realised is the VAT system. Currently the CSO's Business Register uses this system as a primary source for information on the births of enterprises. However, in most EU Member States, the VAT system is also one of the main sources of statistics on the turnover of enterprises. In Ireland, the information on turnover in the VAT system is incomplete because there is a large zero-rated sector. An additional difficulty is that some Traders' Annual VAT 3 paper returns are scanned and held as images. The data are not captured as individual variables and cannot be automatically processed for statistical analysis.

Because of the limited time available on this project, the SPAR team could not pursue these issues. A more detailed analysis will therefore have to be undertaken later when the NSB's initiative on social statistics is extended to the economic and environmental statistics areas.

7.7 Statistical potential

Income Tax file

Much of the data available in the income tax system is already analysed in detail by the Statistics Branch of Revenue. However, this analysis primarily concentrates on those items of most interest in the administration of the tax system. The SPAR team concluded that, by utilising additional information that can be derived from the existing system, the present analyses could be extended to provide some additional useful information on the distribution of incomes and the structure of employment. However, it is recognised that these additional analyses cannot be produced easily and would require rewrites of existing IT systems.

Some key social demographic variables are not currently captured in the tax system and it is therefore not possible to use the Revenue Commissioners' data sets to generate additional data such as an analysis of incomes classified by age. This would be possible if the PPS Number were used to link the tax data to the CRS system in DSFA. Income could also be classified by other key socio-economic variables identified on the CRS.

ST21 form

The usefulness of this information for social statistics appears limited. However, as described above, the source is a good comprehensive source of information on transfers of real property and, if more extensively processed, it could well provide useful support material for economic statistics. In fact, the ST21 details are also supplied to the Valuation Office, which processes them and builds up a comprehensive profile of land transfers coded by geographic region. This information is currently used by the CSO as the data source for its quarterly publication on agricultural land sales.

Inheritance/Gift Tax

Given the high threshold and the existence of various reliefs and exemptions, it is unlikely that this data set would provide useful social statistics.

Forms CA24 and CA4

Again, the incomplete coverage of both of these forms means they are of little interest for social statistics.

7.8 Conclusions

Income distribution statistics

Information on income levels is most useful when results are analysed by social and demographic variables such as gender, age, regions and household compositions. However, the information collected in the tax system is essentially the core data needed to administer the tax systems and such classifiers are often not available. It seems impractical to try to extend the tax reports to collect additional information. This would be expensive and difficult and might ultimately be unsuccessful. A better and more cost effective option would be to extract the required classification variables from elsewhere. The statistical potential of Revenue's income statistics would be greatly enhanced if linkages could be developed to the DSFA's Central Records System (CRS). As described in Chapter 6, the CRS contains demographic and socio-economic descriptors and these could be imported onto the Revenue records using the PPS Number that is common to the two files. Cross-departmental linkages like this are discussed further in Chapter 8 and may take some time to develop.

In the meanwhile, it is worth considering whether the information provided in the existing Income Distribution tables in Revenue's Statistical Report could be enhanced based on the information currently available in the tax system. Despite its shortcomings, the income tax system is still the most comprehensive annual sources of information on incomes in the State and should be analysed to the greatest extent possible. The existing analyses published by Revenue show the numbers of tax cases in different income bands, classified by type of taxpayer. Jointly assessed married couples with both spouses working are recorded as a single tax case and the income shown is the combined income of the two partners. Information in the basic tax system (but not in the combined data set used for existing statistical analyses) may make it possible to separate out the gross income and sex

of the individual partners. If this information is not always currently available, Revenue should consider the insertion of “husband” and “wife” labels on the Income Tax returns in order to improve gender-based statistics. This would allow classification of individual incomes by gender and provide useful additional information on the relative incomes of men and women. It should be noted that such an analysis is only meaningful for pre-tax incomes. In the case of jointly assessed married couples, the couple determines the amount of tax credits allocated to each partner and it is not sensible to look at the net of tax income of individual partners.

Possibilities for other useful cross-classifications are limited. One could produce analyses of incomes for certain sub-groups such as lone parents and retired persons, since such markers are on the tax files. However, the majority of these will be low-income cases and, as mentioned earlier, their coverage on the tax files may not be comprehensive or up to date. This seriously diminishes the usefulness of the existing data for income distribution analyses.

The long-term goal should therefore be to produce regular and more comprehensive reports on the distribution of the after-tax incomes of families and households, by enhancing the information in the tax system with data from other sources. As explained earlier, the existing information on incomes in the income tax system is incomplete. Certain types of incomes and low-income families are not always captured in the system. Incomes are also only available for tax cases (single or married), while income distribution analysis is most concerned with the composition and combined incomes of families and households. Any meaningful analyses would have to include low-income households, whose incomes may not be subject to tax and it is proposed that information on these should be also obtained through linkages with DSFA’s data holdings. Their systems can provide information on social welfare payments by PPS Number and this information could be merged with the tax records to provide a much more comprehensive picture of incomes than currently exists. This, however, requires that the two departments classify social welfare payments in a way that is recognisable by both systems, given that some are subject to income tax and will already be captured as income in the tax files. The analysis also needs to be undertaken at the level of the family or household and this is a more difficult issue. However, the DSFA’s PRETA scheme already captures information on all persons in the household regardless of their dependency status and one of the recommendations in this report is that this information should be collected in all social welfare schemes (see Chapter 6). This would allow individuals, identified by PPS Number, to be combined into households and data on incomes to be aggregated to give overall household incomes. National experts in the field of income distribution analysis should be consulted before detailed specifications are drawn up for the production of the enhanced income statistics.

Structure of employment

Additional microanalysis of the information provided in the P35 forms returned by employers could provide additional useful information on the structure and nature of employment. However, any analysis will be somewhat limited without an “hours worked” indicator. The inclusion of such an indicator would be desirable but is likely to create an unacceptable reporting burden on employers. Nonetheless the P35 return still provides a complete coverage of employees since every employer must complete a return and report for each employee: the PPS Number; the taxable income; and the dates of commencement and termination of employment. Using PPS Number classified by employer number and dates of employment, it should be possible to statistically analyse movements from one employment to another (PPS Number captured leaving one employment and subsequently appearing on another employer’s return). Cases of multiple employments (same PPS Number in two or more P35 returns and covering the same employment period) could also be identified and

linked to the levels of income, although as mentioned earlier, the usefulness of such information is diminished by the absence of data on hours worked.

When analysing the P35 data system, care will have to be taken to separately identify records relating to the payment of pensions to retired persons. Some occupational pensions are within the scope of the PAYE system and the various pension schemes must submit P35 returns. The numbers of such schemes are limited and, if necessary, they could be manually identified and separately processed. Existing data on occupational pension schemes is very limited and the information on these schemes obtainable from the P35 system would in itself be valuable.

The macro-analysis of employment and income statistics will also be greatly enhanced when the activity coding of Revenue records is updated in line with the new international activity classifications. The Director General of the CSO has agreed that the CSO will provide assistance in this area and the necessary work is already underway.

Incomes by regions

The SGSES report identified considerable demand for information on incomes classified by regions. After examining the income tax system, the SPAR team is of the view that some place of work and place of residence linkages are possible by matching the Form P35 employer address file with the employee address file. However, the usefulness of this analysis is limited. All employees of a company are usually included in a single P35 return. These employees may work in separate locations in different branches of the company, but the P35 return may record only a single address designated for tax purposes. The location of the actual place of work cannot therefore be identified in the tax system.

ST21 form

The ST21 captures details of house sales not involving a mortgage. It therefore has the capacity to provide information on house prices that would complement the information already available to the DoEHLG from mortgage lenders. Information from the ST21 form could also provide useful statistics on new building and construction projects. However, while the ST21 system comprehensively records property transfers, it is not an exhaustive data source for new residential, commercial or industrial developments. At present, the form distinguishes between new and second-hand property only in the case of dwellings. This distinction would also need to be introduced for industrial and commercial properties. Even then, certain types of new developments would not be captured. For instance, if a development company buys land for building purposes, the transfer will be recorded on the ST21 return. If the company then builds an office block or shopping there is no requirement to complete another ST21 form regarding the new building unless it is actually sold or let under a lease of 30 years or more. Extensions of existing buildings or new buildings developed on the grounds of existing buildings may also be excluded. Therefore, the ST21 system will never be capable of providing comprehensive information on new building output but it has the potential to be a very useful check for verifying the comprehensiveness of construction statistics.

The SPAR team therefore recommends that the ST21 form be expanded to separately identify new properties also in the case of commercial and industrial developments. The team also recommends that the CSO liaise with the Valuation Office to determine if the existing information on agricultural land sales, which is partly based on the ST21 information, could be expanded to provide information on the sales of other properties.

7.9 Summary recommendations

(i) Develop a new set of annual employment statistics based on the existing P35 income tax returns.

(ii) Investigate the possibility of separately identifying the gross incomes of husbands and wives in order to produce additional breakdowns of income by gender in the Income Distribution tables published in Revenue's Annual Statistical Report.

(iii) Expand the property breakdown in the ST 21 form to separately distinguish transfers of new commercial and industrial properties, with a view to providing useful data for building and construction statistics. The CSO should also liaise with the Valuation Office to determine if the information they currently provide on agricultural land sales can be extended to cover information on the sales of other properties.

(iv) Extend the coverage of incomes on the Revenue files by merging information on non-taxable incomes from the DSFA data holdings. In the longer term, the possibility of compiling annual household income estimates using individual-household linkages collected in DSFA schemes should be considered.

Chapter 8

Cross-cutting Statistical Issues

8.1 Introduction

This chapter examines a number of issues that would simplify taking a cross-cutting approach to policy-making across the six departments examined in this report. Progress on these issues would clarify the inter-relationships between policy questions in different departments. These issues are such that they can best be progressed through high-level inter-departmental co-operation. They include:

- The use, for statistical purposes, of a common personal reference number (the PPS Number) for all transactions between individuals and government departments (see Section 8.2);
- The use of address postcodes or geo-reference co-ordinates to assist in the compilation of small area statistics (see Section 8.3);
- The use of and sharing of expertise on standard classifications and coding systems (see Section 8.4); and
- Statistical management practices that increase the statistical potential of the data holdings (see Section 8.5).

The chapter ends with a summary of the typical statistical strengths and weaknesses found in administrative records, and with a brief review of the situation with respect to surveys that are funded by government departments and undertaken, on their behalf, by outside bodies.

8.2 Data integration

The PPS Number was introduced in the 1998 Social Welfare Act as the unique personal identifier for transactions between individuals and government departments and agencies designated in the Social Welfare Acts. As indicated in Section 6.8.1, these include key departments for social statistics such as Education and Science; Environment, Heritage and Local Government; Health and Children; and the Office of the Revenue Commissioners. Legislation provides that only designated public bodies may use the PPS Number. The CSO is a designated body. The SPAR teams found that the PPS Number was being used in 63 of the data holdings examined in this report (see Appendix 3).

The Central Records System (CRS) file maintained by the DSFA is a register of all PPS Numbers. The CRS holds key demographic and socio-economic information of interest for policy purposes, such as sex, date of birth, date of marriage, marital status, nationality, household ID number, income data, occupation and employer details.

Apart from its administrative use, the CRS has the potential to be of enormous statistical benefit. The CRS is already being used by the CSO to reduce the statistical burden on respondents in household surveys. For example, the new CSO survey on Statistics of Income and Living Conditions (EU-SILC) asks interviewees to allow the CSO to use their PPS Number to obtain financial information directly from DSFA and Revenue rather than directly during the interview. The PPS Number is also being collected in the CSO's new National Employment Survey. From a statistical perspective, this usage reduces the response burden and provides the CSO with more detailed and more accurate information. It also makes more effective use of interviewers' time, allowing the collection of information not available through administrative sources such as perceived health status, and unreported crime.

On a broader level, many issues of importance to society can only be fully understood if information from different data sources can be combined. This requirement applies to both data sources held within the same department and to data sources held in different departments. The use of a common reference number would allow the CSO to increase the statistical value of the separate individual sources, and to support the kind of in-depth analyses needed by policy-makers and business analysts. Significant statistical analyses could be undertaken across public sector databases, resulting in a comprehensive statistical profile of the population in terms of education, health, employment etc., if the individual data sources could be brought together at the person, family and household levels. For example, the progress of individual pupils through the education system could be more thoroughly studied if pupil databases existed that could be integrated across first, second and third level cycles.

Under the Statistics Act, 1993, the CSO is responsible for ensuring that the statistical potential of the records maintained by public authorities is realised. Section 30 of the Act gives the CSO right of access[26] to the data holdings held by, or funded by, government departments. Maximising the potential and value of administrative data will require the CSO to integrate administrative data files. Such integration would allow the provision of a comprehensive demographic and socio-economic background to the key Irish social statistics. The SGSES report also acknowledged the essential importance of data integration and recommended an approach to allay any fears the public may have about the preservation of the data protection rights of individuals. The NSB proposed that:

- The CSO be asked to set out formally how its process of data integration and the subsequent treatment of statistics generated by data integration (including access for research) can be safely employed without data protection problems;
- The NSB in conjunction with the SOGSI should be asked to set out formally how departments would use and protect individual data available to them for statistical purposes; and
- These documents should be referred to the Data Protection Commissioner for confirmation that this process does not undermine the data protection rights of individuals.

The SPAR teams endorse this approach.

8.3 Spatial aspects

The Government published the National Spatial Strategy 2002-2020 (NSS) in May 2003. The NSS is a planning framework designed to achieve a better balance of social, economic and physical development, and population growth between regions. The high growth rates achieved in Ireland during the 1990s have altered the socio-economic geography of Ireland. It is essential that data holdings in departments should contain detailed and consistent geographical identifiers to enable policy-makers to accurately measure these regional changes. We are also aware that in some countries (e.g. the UK) there has been significant investment in developing neighbourhood statistics to support “joined-up” policy responses. This section briefly examines three recent developments in this area.

[26] The Act does not give the CSO automatic right of access to records pertaining to a Court, the Garda Síochána or the prison administration. The agreement of the Minister for Health and Children is required for access to medical records that are not publicly available.

Irish Spatial Data Infrastructure

The Government's Action Plan for developing an Information Society, *New Connections*, was published in March 2002. It included plans for a national Spatial Data Infrastructure (SDI) to be progressed as a strategic priority in the context of the overall development of the e-Government process. The DoEHLG was given the central role in establishing the SDI.

In essence, a SDI is an initiative to ensure that spatial data are organised in ways that more effectively describe the national economy, promote social inclusion, meet citizens' needs, deliver services more effectively, and help protect or enhance the environment.

In SDI, a distinction is made between reference data (direct information for a position in space such as the course of a river or an administrative boundary), and statistical data, such as population, that can be related to position in space. A SDI is not primarily concerned with the actual statistical information but only with the way in which it is linked or 'tagged' to spatial location. There is a need to ensure that the reference information and the tagging are organised so as to achieve the benefits associated with the coherent organisation of spatial information.

The SDI is an important cross-cutting initiative that should increase the potential of administrative data, particularly if proper standards are agreed and adhered to. The DoEHLG has established an ISDI (Irish SDI) Work Group, comprising Geographical Information Systems and spatial experts. The CSO became a member of the ISDI Work Group in July 2003. This group will report to the National Spatial Strategy Implementation Committee. At its inaugural meeting earlier this year, the Work Group agreed a two-pronged approach:

- To determine user needs in this area; and
- To undertake a scoping study to identify which organisations generate and hold significant amounts of spatial data, and to obtain details regarding key features of the data (metadata, accuracy, format, frequency of updating, etc.).

There are similarities between the ISDI and SGSES/SPAR initiatives to the extent that both are concerned with data availability, data needs, and standards to improve the accessibility and usability of the information. Indeed, the DoEHLG have indicated that they consider it necessary for them to examine in detail the findings of the current SPAR Project before undertaking any research initiative in the area of SDI.

Although the Spatial Data Initiative is still at a very formative stage, the SPAR teams strongly recommend that all relevant organisations be fully aware of the needs of the emerging SDI initiative and take it into account when organising the spatial elements of their data.

GeoDirectory

The GeoDirectory is a collaboration between An Post and Ordnance Survey Ireland (OS*i*). The purpose of the directory is to create a definitive reference directory of addresses in Ireland and to assign them accurate postal and geographical addresses. To this end, An Post and OS*i* propose using a standard format for all addresses in Ireland. The directory includes a geographical address (grid reference co-ordinates) but it does not include a Postcode.

Postcodes

The SPAR teams support the introduction of an addressing system that would unambiguously identify the location of an address independently of the address description. ComReg published a consultation paper on the Regulation of Postal Services in November 2002[27]. The paper invited opinions on a number of issues including the possible introduction of postcodes. Both the CSO and the NSB responded with opinions in favour of their introduction. Standardised postcodes would facilitate the integration of results from separate surveys at a small geographical level in a much more exact way than attempting to derive geographical area information from text addresses or grid reference co-ordinates. The ISDI working group should examine and make recommendations on the need for postcodes.

8.4 Data classifications and coding practices

The processing of data from a survey or administrative scheme generally requires the use of data classifications and coding. For example, employment schemes need to record the previous occupations of applicants. To analyse these in an efficient and effective manner, it is necessary to adopt a consistent level of detail for recording a person's occupation and for subsequently coding that information in a computerised format.

In the late 1990s, the CSO set up an office-wide Classification and Standards Section (CARS) with responsibility for this the establishment of classifications standards and the co-ordination of their use within the CSO (see Appendix 9). The CARS Section has to-date primarily focussed on itemising and standardising the coding systems in use within the CSO. They have also recently begun consulting with statisticians in other government departments. The SPAR teams recommend that this practice be formalised and that CARS section should take a more proactive role in ensuring that standard coding systems are being used throughout the public service. For many variables, it should be sufficient for the CSO to make available to departments a recommended standard classification/coding system. However, for a complex variable such as occupation, it may be necessary for the CSO to take a more active role such as developing and sharing software that facilitates automated coding of text descriptions. Using these tools, it is possible to make considerable time and cost savings in such coding.

Table 8.1 examines the classification information returned to the CSO in connection with the data sources examined by the SPAR teams (see Appendix 3). The table shows that there is a wide variety of coding systems being used under each classification variable. In some cases, it may be possible to establish one-to-one relationships between some of these codes. However even that may not be possible over time unless those coding systems are being kept up to date.

Age coding information was received in relation to 65 of the data sources that were reviewed by the SPAR teams[28]. While there were nine different age coding systems in use (see Appendix 5), 51 of the 65 coding systems were consistent with CSO best practice. A less favourable situation exists for nationality. There were nine different nationality-coding systems in use. Of the 37 sources where nationality-coding details were provided, only two were using what would be regarded by the CSO as a standard classification. Given the importance of immigration to understanding recent social change in Ireland, there is a clear need to be able to reliably classify information by nationality. An example of the need for such information is in CSO population projections - the models behind these projections

[27] ODTR (2002): ODTR 02/95 - Regulation of Postal Services.

[28] Appendix 3 contains a list of the data sources reviewed by the SPAR teams. Age information is not necessarily collected in each of these data sources.

would benefit from the availability of accurate information on the birth rate of different nationalities.

Table 8.1 Analysis of selected classification variables

Classification variable	Data sources collecting variable	Data sources where coding details were obtained	Forms of coding identified	Data sources using CSO / other standard coding[29]
Age	93	65	9	51
Gender	107	69	18	0
Marital status	77	61	20	0
Socio-economic status	72	57	17	3
Income	67	59	10	0
Geography	110	92	16	3
Phone number	57	56	5	0
Family status	61	44	11	0
Disability	48	36	13	0
Nationality	80	37	9	2
Race/Ethnicity	12	5	5	1
Membership of Traveller community	15	8	6	0
Sexual orientation	3	2	2	0
Religious affiliation	11	4	4	0
Farm/Non-farm household	31	22	7	0

Appendix 8 summarises the actual coding systems being used for nationality. Table 8.2 lists the nationality codes being used for the EU member states and the 10 countries due to join the EU during 2004. The wide variety of coding systems, and the fact that some countries are not coded, limits the statistical potential of these data sources. The ISO 3166 coding level of detail is essential for pivotal systems, such as birth registration, where the nationality of the parents is recorded.

Use of standard classifications across data collection systems makes it possible to compile consistently categorised results for all information sources. For example, the use of common age groups such as five-yearly bands (0-4, 5-9, etc.), or common geographical levels such as county, more readily allow users to build up profiles of particular age groups or local regions. Hence, the issue of standardised coding relates not only to using common coding systems at the individual record level (e.g. date of birth) but also common or consistent age groups at the macro level for dissemination. CSO can provide a standard methodology and toolkit to government departments to facilitate central storage, querying of macro data and subsequent publication of such data. However, the integration of macro data can only be realised if standard dissemination classifications are being used across government departments.

[29] Figure for Age refers to the number of data sources which collect date of birth.

Table 8.2 Examples of nationality codes used for EU and accession states

Country	ISO 3166 / CSO Standard[30]	DSFA – ISTS	DSFA - PENLIVE	GRO	DES Post-Primary Pupil Database
Belgium	BE	BEL	107	21	56
Denmark	DK	DEN	109	56	208
Germany	DE	GER	156	79	276
Greece	GR	GRE	110	82	300
Spain	ES	SPA	112	198	724
France	FR	FRA	105	72	250
Ireland	IE	IRL	0	103	372
Italy	IT	ITA	108	105	380
Luxembourg	LU	LUX	111	125	442
Netherlands	NL	NET	106	152	528
Austria	AT	AUS	115	14	40
Portugal	PT	POR	113	172	620
Finland	FI	FIN	125	71	246
Sweden	SE	SWE	123	204	752
United Kingdom	GB	UK	157	223	826
Cyprus	CY	CYP	146	54	196
Czech Republic	CZ	-	-	55	202
Estonia	EE	-	-	66	228
Hungary	HU	-	150	95	348
Latvia	LV	-	-	118	420
Lithuania	LT	-	-	124	440
Malta	MT	-	122	133	470
Poland	PL	POL	139	171	616
Slovakia	SK	-	-	192	672
Slovenia	SI	-	-	193	676

8.5 Statistical management of data sources

To ensure that the statistical potential of the data holdings is fully realised, it is essential that staff responsible for these data sources are aware of best practice in maintaining registers, processing the completed forms, and disseminating the results. Statistical management encompasses issues such as:

- Register procedures that clearly distinguish between current and inactive records;
- Capturing the raw data once and at the point closest to the data provider;
- Data capturing practices that are consistent with recognised international statistical classifications and standards;

[30] The CSO standard classification of countries is based on the ISO-3166-1:1997, but includes a hierarchical element, which accommodates the constituent parts of the United Kingdom.

- Procedures that clearly document treatment of missing and suspect data; and
- Management of the IT aspects of statistical systems in an integrated and clearly documented manner.

The SPAR teams encountered many problems relating to the management of registers that limited the statistical potential of the associated data sources. These included problems with how to handle exits from the register, recording the timing of events such as new entries, or amendments to existing entries such as updating educational qualifications. These difficulties are particularly limiting if such registers are being used for selecting samples or for grossing the results of sample surveys to the population. The inclusion of key additional variables on a register can also result in selecting more efficient and smaller samples.

Examples of IT management of statistical systems that are consistent with facilitating data integration are evident in a number of departments. For example, DSFA has reduced the variety of their IT systems by grouping their individual schemes into a small number of systems.

The Integrated Housing Project currently being developed in DoEHLG is bringing together 16 previously independent systems into one common system. The project involves standardising the application forms across these schemes and capturing a considerable amount of applicant information that was previously only available on paper.

The new DoEHLG House Price System represents another example of good practice. At present, various institutions compile price trend reports based on potentially non-representative subsets of house purchase data. Under the new system, the department will receive an electronic file from the mortgage institutions of the individual mortgage transactions in an anonymised form. The new system should result in cost savings (data being captured once), improved timeliness and extra detail on the borrower and on the property. This will allow more comprehensive statistical analyses to be compiled including a more exact distinction of the price trends for different types of properties (e.g. new/second-hand, location, type, size, and period of construction).

The redesign of IT systems or the establishment of new data holdings provide ideal opportunities for departments to review the quality of their statistical management practices. Recognising this, the NSB has asked the CSO to take a lead role, in line with its remit under the Statistics Act, 1993 in developing a collaborative network among data producers[31]:

> *"This should involve all government departments and agencies that hold administrative data capable of generating worthwhile statistics. An immediate consequence is that departments and agencies should consult with the CSO when they are undertaking new data collection or changing information technology systems. It is easiest to maximise the statistical potential of administrative records if this is built in at the design stage of IT systems."*

Given the CSO experience in these issues, sections in departments and agencies administering data sources should, as a matter of course, seek advice from the CSO on good practice in the statistical management of data holdings. Although not discussed here in detail, such practices also include a focus on completeness and accuracy of data capture, data editing processes for checking and correcting data, and estimating for missing information. They also include practices that ensure all procedures and data holdings are

[31] NSB (2003), "Strategy for Statistics 2003-2008"

fully documented. Indeed, the CSO is currently undertaking a major exercise to improve the quality of its survey metadata so that all relevant documentation is attached to, or in close proximity, to the data file itself.

8.6 Typical strengths in administrative records

In recent years, there has been a significant increase in the extent to which administrative records have been computerised and in the coverage of administrative schemes. The CSO recognises the benefit in optimising the use of existing administrative data holdings before formulating proposals for new surveys to meet the increasing national and international demands for statistics. Records available through administrative sources have a number of potential advantages over attempting to collect the same information by means of a statistical survey. Some of these advantages are:

- The cost is often significantly lower as the statistics are collected as an add-on to an existing administrative scheme;
- Administrative schemes often have a higher and more timely response particularly if there is an administrative obligation or an economic incentive to participate;
- Such schemes can cover specific sub-groups of the population in a more effective way than is feasible using general household sample surveys;
- It is easier to collect financial information; and
- The workload is often spread across local offices.

8.7 Typical statistical weaknesses in administrative records

Throughout Chapters 2-7, a number of recurring weaknesses have been identified that tend to limit the potential use of administrative records for statistical purposes. These weaknesses include:

- Administrative systems are often application or case-based rather than person based;
- In some departments, data capture and processing is very fragmented and statistical procedures are often not applied consistently at local level;
- For a number of data holdings, only aggregated information is available. This prevents the undertaking of more thorough analyses and longitudinal studies;
- There is an increasing tendency to use the PPS Number as a unique reference number. However, increased use of this identifier is subject to meeting data protection requirements;
- Without widespread use of a common personal identifier, there is no ready mechanism to combine individual related data holdings for cross-cutting analyses either within or across departments;

- There is repetitive data collection across and within departments. In some cases, departments are duplicating compilation of client registers with very significant overlap e.g. CRS, GMSPB, Electoral register;
- There may be a lack of central awareness within a department of all its data holdings;
- It is difficult to aggregate administrative records at the individual level to larger units such as the family or household;
- There is a lack of usage of standard classifications and coding systems;
- Administrative forms are sometimes only partially computerised;
- The publication of statistical information from administrative schemes should be within regular time deadlines;
- The voluntary nature of many schemes often means that all eligible applicants do not apply;
- The absence of an agreed core set of demographic and socio-economic variable, for inclusion in each scheme, has reduced the statistical value of each data holding.
- While it is important to retain inactive records in administrative registers, there needs to be a clear method of marking critical events such as the time at which a record first became active or ceased to be active; and
- There are difficulties in updating useful statistical information, such as educational qualifications, where such data are not essential to the administration of a scheme.

8.8 Access to surveys funded by government departments

The NSB called for a broader and more collaborative approach to the collection of public service statistics in its Strategy for Statistics 2003-2008. The NSB asked the CSO to examine how the statistical potential of administrative records and department-funded surveys could be better integrated into official statistics to meet the increasing needs of users. Throughout this report, there have been various instances of statistical work being undertaken by outside organisations on behalf of departments.

Table 8.3 outlines some of the key government funded surveys undertaken on behalf of government departments. The method of public dissemination for these surveys varies from a standard statistical report to the irregular release of specialised analytical reports. These data sources have a broad societal and policy interest. Standard statistical reports should be published from publicly funded surveys and the funding department should be the ultimate owner of, and have full access to, the raw data. The NSB also expressed the view (Section 3.3 of Strategy for Statistics 2003-2008) that data from all such surveys conducted at the taxpayers expense are automatically part of Ireland's official statistics and must be accessible by the CSO.

Table 8.3 Some examples of outsourced Department funded social surveys

Department	Survey
Education	Annual School Leavers Survey
	International Adult Literacy Survey
	Program for International Student Assessment
	Surveys of achievement at Primary and Secondary level
	National Assessment of English Reading
	National Assessment of Mathematics Achievement
	Suirbhé Náisiúnta Inniúlachta sa Ghaeilge I Rang a Sé
	Survey of Reading Literacy in Designated Disadvantaged Schools
Environment	National Survey of Housing Quality
Health	National Perinatal Reporting System
	Hospital In-patient Enquiry
Justice	EU Migration Network
Social and Family Affairs	Living in Ireland Survey
Other	European Social Survey
	National Survey of Domestic Abuse

Chapter 9

Conclusions and Recommendations

9.1 Introduction

Some detailed conclusions and recommendations have already been presented in Chapters 2-8. This chapter focuses principally on the general conclusions and key recommendations that could have significant outcomes for a specific data holding or for a cross-cutting issue. The chapter also includes a brief review of the policy context driving an increased demand for official statistics.

9.2 A brief review of the policy context

In recent years, there has been an increasing demand from society and Government for statistical evidence on which to base and evaluate policy. The Government articulated its social objectives in a range of reports such as the Social Partnership Agreements, the National Anti-Poverty Strategy (NAPS), the National Children's Strategy[32] (NCS) and the National Spatial Strategy. The NAPS recognised that poverty is mainly concentrated in households with no person in employment or only persons in low-paid employment. These include households with elderly people, people with disabilities, carers, one-parent families, larger families and minority groups. The NCS recognised the need for consistent, timely and relevant information on children and concluded that a national longitudinal study of children was needed to correct the existing information deficit.

The new demands were partially met by the CSO developing the QNHS as a means of collecting a broad range of information on social issues through the incorporation of special survey modules on topics such as crime, pensions, disability and childcare. However, the NSB and CSO realised that a more fundamental structural approach was needed to identify and meet the key needs of users. The NSB reviewed these developments in depth during the development of its new strategy for statistics. It concluded that the new demands for information were such that a new statistics network needed to be developed across departments. It was envisaged that this network would work in a collaborative manner to ensure that the production of the new statistics being demanded by society and Government would be imbedded into the system. Essentially, the challenge to the statistical system has become one of anticipating future statistical needs and establishing a framework that is comprehensive and flexible enough to deliver them. This Report has attempted to identify some of the practical steps involved, and some of the value that might be unlocked, in implementing this whole-system approach.

9.3 Conclusions and recommendations

The SPAR teams are satisfied that all of the major social data sources in the six departments have been reviewed and that these are sufficiently representative for the teams to draw useful conclusions and recommendations. The data sources contain a vast amount of information on topics of central interest to Irish society. However, in most cases, they are contained in stand-alone information systems primarily designed to meet administrative requirements rather than the needs of policy-makers. While there are significant ongoing efforts on the part of data holders to improve the quality of the statistical information being derived from these sources, these efforts need to be organised in a more structured and strategic framework.

Many of the recommendations made by the SPAR teams are with the intention of adapting the data sources to optimise their statistical potential. Development of this potential will also help departments to meet their own internal data needs. This should be of use to

[32] DHC (2000): "National Children's Strategy"

departments when they are formulating the data/statistics strategies requested by the Government in response to the SGSES memorandum. The teams found that there was no ready repository of summary information on all of the data holdings being held within a department and its agencies. In addition, departments did not always have access to the full range of information held by agencies and other department bodies.

The SPAR teams encountered recurring problems relating to the management of registers that limited the statistical potential of the associated data sources. These difficulties reduce the potential value of the registers for maintaining up-to-date and comprehensive information and for selecting or grossing survey samples if required. In addition, the inclusion of key additional variables on a register would facilitate the selection of more efficient and smaller survey samples.

The data protection laws in Ireland strictly limit the potential use of the PPS Number for sharing of administrative information between departments. While these limitations protect the rights to privacy of citizens, they may also conceivably operate to the detriment of individual citizens in certain circumstances. Furthermore, they also result in a higher burden of administration and repetitive form filling than would be the case if the statistical aspects of this information could be shared between government departments. For example, there would be significant cost savings and statistical benefits if the key demographic and socio-economic information on all persons living in Ireland could be collected centrally and shared both within and between departments under strict legal agreements. This would avoid the current situation where departments for different purposes are separately duplicating effort and response burden in constructing their own partial population registers.

A culture of good practice in managing data holdings has developed over time in the CSO. The teams consider that it would be of significant benefit to departments if they also had this expertise available to them either internally or through the CSO. Increasing the collaboration between producers of official statistics is one of the key priorities in the NSB's strategy 2003-2008. The CSO has recently placed a number of statisticians in government departments and this approach could, over time, contribute to developing the statistical competency of producers and users of statistical data in those departments. However, for these staff to be fully effective within these departments and as part of the broader public service statistical system, they need to operate as a central statistical unit and have a key input into standardising the collection, processing and dissemination of each department's data holdings. In addition, such units would need to work closely with the CSO. These changes are needed to ensure that consistent standards are in use throughout the administrative data process which ranges from the initial conceptualisation of the information required through to designing application forms or survey questionnaires, specifying IT procedures for capture and editing of data, integration of data at the micro level, and the dissemination of results. Only then can departments be confident that they will have the cross-sectional and longitudinal statistics that will enable them to fully understand the key determining factors influencing social, health and educational outcomes.

General recommendations

These recommendations are of a cross-cutting nature largely directed towards the development of an integrated and comprehensive approach to the production of statistics across the public service. Implementation of these recommendations will greatly facilitate the undertaking of cross-sectional and longitudinal analyses that explore more fully the complexity of society in Ireland.

Recommendation 1: Departments should ensure that related data sources are capable of being integrated at the individual record level.

Recommendation 2: There should be greater collaboration between government departments and the CSO to ensure that the statistical potential of administrative records is realised efficiently, and that duplication within the system is minimised.

Recommendation 3: Departments should establish mechanisms to regularly review their information needs as the statistical information available from official data sources develops.

Recommendation 4: Some social statistics are more meaningful at the individual, family or household level. Where relevant, data holdings should be structured so that case-level or event-based information can be aggregated to compile individual, household and family statistics.

Recommendation 5: All official data sources should use standard classifications and coding-systems. The CSO should take a lead role in promoting, maintaining and disseminating these classifications.

Recommendation 6: The CSO should work with departments to develop a core set of demographic and socio-economic variables. These could either be independently collected in administrative schemes and surveys or preferably, subject to meeting data protection restrictions, available via a central repository such as the CRS. Spatial information and nationality should be included in this core set.

Recommendation 7: Individual departments should develop a common look and content in the design of their administrative scheme application forms. This will ensure that the core information is collected in an identical manner across all schemes.

Recommendation 8: Departments should give priority to developing the data holdings that have information on persons not living in permanent households, i.e. the homeless and persons living in institutions, as these are outside the coverage of the CSO sample household surveys.

Recommendation 9: All statistics compiled in government departments should adhere to the UN Fundamental Principles of Official Statistics (see Appendix 10). In addition, where statistical reports are published from department data holdings, these should be published in a consistent format and adhere to a defined publication timeliness.

Recommendation 10: All departments should be aware of the emerging SDI initiative and take it into account when organising the spatial elements of their data. The ISDI should examine the policy needs for postcodes.

Recommendation 11: Departments should increase the level of statistical competency in data processing and analysis among their staff.

Recommendation 12: Departments should ensure that all statistical information of use to policy-making, collected in administrative schemes, is computerised.

Specific recommendations

Although each of these recommendations is specific to an individual department, implementation of them would have cross-cutting benefits in many cases.

Recommendation 13: DES should take measures to develop an integrated individual pupil database covering first, second and third level cycles of education. The PPS Number should be included in the State examination databases.

Recommendation 14: The inclusion of the PPS Number in the Register of Electors should be considered.

Recommendation 15: The computerisation of the GRO should be completed as soon as possible, and the GRO should resume providing CSO with the data required for disseminating detailed statistics on marriages.

Recommendation 16: DHC should review the methodology of the Hospitals Waiting Lists, and in particular the definition of waiting time. The data should be captured at an individual level to allow analyses by age, gender, average waiting time, and other key characteristics. The DHC should examine whether waiting times could be captured in the HIPE system.

Recommendation 17: DJELR should urgently address the statistical quality and access issues identified with the Garda PULSE System (see Section 5.2).

Recommendation 18: Statistical information relating to the issue of visas should be computerised. This information could give an early indication of future immigration patterns.

Recommendation 19: The DSFA should give priority to addressing the statistical limitations of the CRS, with respect to local area statistics and full coverage, in order to improve its register value for statistical purposes.

Recommendation 20: Revenue should agree a mechanism with DSFA which would enable them to extend the coverage of low income earners in their income tax statistics and to present these statistics according to the key demographic and socio-economic variables.

Recommendation 21: The CSO and Revenue should jointly examine how the P35 records could be used to produce additional employment statistics.

Appendices

Appendix 1 Membership of CSO teams

Focus Department	CSO team	CSO support team
Education and Science	**Ger Healy** Diarmuid Reidy Gillian Wall	Steve MacFeely John O'Leary
Environment, Heritage and Local Government	**Adrian Redmond** Jim Dunne John Dunne	Gerry Walker Joe Madden
Health and Children	**John Fitzpatrick** Gordon Cavanagh Niall O'Hanlon	Paddy McDonald Kevin McCormack
Justice, Equality & Law Reform	**Aidan Punch** Mark Davis Paul J Crowley	John O'Hagan Liam Hogan
Office of the Revenue Commissioners	**Mick Lucey** Máire O'Mahony Kevin Phelan	Pat Fanning Elaine Lucey
Social and Family Affairs	**Padraig Dalton** Deirdre Cullen Richard McMahon	Catherine Finneran Tom McMahon
CSO support team	**Gerry Brady** Denis Murphy Patrick Quill Gillian Roche	

Appendix 2 Reference documents

Author(s)	Document
Department of Education and Science	Statistical Report 2000/2001
Department of the Environment, Heritage and Local Government	Report of the Commission on the Private Rented Residential Sector
Department of Health and Children	Health Strategy: Quality and Fairness: A Health System for You
	Report of the Commission on Financial Management and Control Systems in the Health Service (Brennan Report)
	The Health Service Reform Programme
	Annual Report of Breast Check 2001-2002
	The National Health & Lifestyle Survey
	Health Statistics 2002
	Long-Stay Activity Statistics 2001
	National Children's Strategy
Department of Justice, Equality and Law Reform	Strategy Statements 1998-2000 and 2002-2004
	Annual Report 2000
	Report of the Working Group on the Review of the Parental Leave Act 1998
Office of the Revenue Commissioners	Statistical Report 2001
Department of the Taoiseach	Sustaining Progress, Social Partnership Agreement 2003-2005
	National Spatial Strategy
	National Anti-Poverty Strategy
An Garda Síochána	Annual Report of An Garda Síochána 2001
An Post, Ordnance Survey Ireland	www.Geodirectory.ie
Courts Service Board	Courts Service Annual Report 2001
Economic and Social Research Institute	Report on Perinatal Statistics for 1999
	National Survey of House Characteristics and Quality 2001/2002
	Activity in Acute Public Hospitals (1990-1999), Casemix Statistics
Environmental Protection Agency	Environment in Focus 2002
Farrell, Grant, Sparks	Review of Fire Safety and Fire Services in Ireland
General Medical Services (Payments) Board	Annual Report of GMS(P)B 2002
Health Research Board	Annual Report and Accounts 2000
	Annual Reports of the National Intellectual Disability Database Committee
	Activities of Irish Psychiatric Services 2001
Higher Education Authority	Enrolment and Graduand Report 2001-2002

Author(s)	Document
	National Survey of Access to Higher Education
Irish Prison Service	Irish Prison Service Annual Report 2001
Legal Aid Board	Annual Report 2000
National Disease Surveillance Centre	Notifiable Infectious Diseases in Ireland Reports
National Statistics Board	Developing Irish Social and Equality Statistics to meet Policy Needs (SGSES Report)
	Strategy for Statistics 2003-2008
Office of the Data Protection Commissioner	Fourteenth Annual Report of the Data Protection Commissioner, 2002
Prospectus - Watson Wyatt Worldwide	Audit of Structures and Functions in the Health System 2003

Appendix 3 List of data sources examined

Data source title	PPS	Age	Gender	M/S[33]	S/E[33]	Income	G/C[33]	P/N[33]	F/S[33]	Dis-ability	Nation-ality	R/E[33]	T/C[33]	S/O[33]	R/A[33]	Farm
House Price Information proposed to be supplied to Housing Policy and Supply, DoEHLG by Mortgage Lenders - PROPERTIES		X					X									
House Price Information proposed to be supplied to Housing Policy and Supply, DoEHLG by Mortgage Lenders - BORROWERS		X	X	X	X	X	X									X
National House Condition Survey - HOUSES		X				X	X	X	X							X
National House Condition Survey - PERSONS		X	X	X	X	X	X	X	X							X
National Vehicle & Driver File - OWNERS							X			X						
National Vehicle & Driver File - DRIVERS	X	X	X				X	X		X	X					
Register of Electors		X			X		X	X		X	X					
Register of Private Tenancies	X						X									
Proposed Integrated Housing System - PERSONS	X	X	X	X	X	X	X	X	X	X	X		X			
Proposed Integrated Housing System - PROPERTIES					X	X	X			X						
Annual Census of Primary Schools		X	X		X		X			X			X		X	
Annual Returns (Higher Education Colleges)		X	X				X			X	X	X	X			
Annual School Leavers Survey		X	X	X	X	X	X			X	X					X
Annual Statistical Return of students in full-time 3rd level education for non-HEA Institutes		X	X		X		X				X					
Curriculum Provision and Integration among First Year Students		X	X		X		X		X		X		X		X	X
European Social Survey		X	X	X	X	X	X		X	X	X	X	X	X	X	
First Destination Report		X	X		X	X	X				X					
HEA Annual Student Statistics Report		X	X		X		X				X					
International Adult Literacy Survey (IALS, 1994)		X	X	X	X	X	X		X	X	X	X				
Living In Ireland Survey		X	X	X	X	X	X		X	X	X					
National Assessment of English Reading		X	X		X		X			X						
National Contributor to Eurostudent Survey - Social and Economic conditions of student life in Europe		X	X	X	X	X	X			X	X					
National Development Plan In-Career Development Unit element		X	X				X									
National Survey of Access to Higher Education		X	X		X		X									
National Survey of Mathematics Achievement (1999)		X	X		X		X		X	X		X	X			

[33] M/S = Marital Status, S/E = Socio-Economic Status, G/C = Geographical coding, P/N = Phone Number, F/S = Family Status, R/E = Race/Ethnicity, T/C = Membership of the Traveller Community, S/O = Sexual Orientation, R/A = Religious Affiliation

Data source title	PPS	Age	Gender	M/S[33]	S/E[33]	Income	G/C[33]	P/N[33]	F/S[33]	Dis-ability	Nation-ality	R/E[33]	T/C[33]	S/O[33]	R/A[33]	Farm
NCTE Census of Schools					X											
Post-Primary Pupils Database	X	X	X				X				X					
Post-Primary Teacher Payroll	X	X	X				X		X							
Primary Teachers Salary Return Form		X	X	X		X	X									
Programme for International Student Assessment (PISA)		X	X		X		X		X	X	X	X				
Second-level part-time census for vocational, community and comprehensive schools		X	X				X						X			
SOCRATES - ERASMUS in Ireland		X	X				X			X	X					
Staff returns for HEA designated Institutes			X													
Statistical Section, Examinations Branch		X	X							X	X					
Study of Remedial Education in Irish Primary Schools (1997)		X	X							X			X			
Suirbhé Náisiúnta Inniúlachta sa Ghaeilge I Rang a Sé		X	X		X		X		X	X	X	X				
Survey of disadvantage in primary schools		X	X		X		X		X			X	X			
Survey of Reading Literacy in Designated Disadvantaged Schools (2003-2006)		X	X		X		X		X	X	X		X			
Third-level part-time census for IoT's DIT, Tipperary Institute and Killybegs HTC		X	X				X						X			
Annual Survey of Long Stay Units		X	X				X	X								
EUROCAT (ERHA Register of Congenital Anomalies)		X	X	X	X		X		X							
Hospital In-patient Enquiry (HIPE)		X	X	X			X					X	X			
Hospital Waiting Lists		X					X									
Notifiable Infectious Diseases Database	X	X	X	X		X	X	X	X	X	X	X	X	X	X	X
GMSPB Database for Primary Care Reimbursement	X	X	X	X			X	X							X	
National Cancer Registry	X	X	X	X	X		X				X					X
National Intellectual Disability Database		X	X				X	X		X						
National Perinatal Reporting System	X	X	X	X	X		X		X		X					
National Psychiatric In-Patient Reporting System (NPIRS).		X	X	X	X		X			X	X					X
Survey of lifestyles, attitudes and nutrition (SLÁN)		X	X	X	X	X	X			X	X				X	X
National Breast Screening Programme		X	X				X	X		X						
Health behaviours in school-aged children (HBSC)		X	X		X	X	X		X						X	
Civil Registry of Births, Deaths, and Marriages	X	X	X	X	X		X				X					X
Annual report of An Garda Siochana		X					X									
Asylum Management Live System	X	X	X	X	X		X		X	X	X	X			X	
Central Authority for Child Abduction Information		X	X	X							X					
Central Authority for Maintenance Recovery Information		X	X	X		X					X					
Civil Legal Aid		X	X	X		X	X	X	X	X						X

Data source title	PPS	Age	Gender	M/S[33]	S/E[33]	Income	G/C[33]	P/N[33]	F/S[33]	Dis-ability	Nation-ality	R/E[33]	T/C[33]	S/O[33]	R/A[33]	Farm
Criminal Legal Aid								X	X	X	X					X
Equality for Women Measure of the NDP 2000-6	X	X	X	X	X		X		X	X		X	X	X		
Garda Complaints Board		X					X									
Membership of state boards			X													
Prisoner Records System		X	X	X	X		X	X	X	X	X	X	X		X	
PULSE System	X	X	X	X			X				X				X	
Reception and Integration Agency databases - DASS and LOCATIONS		X	X	X		X	X		X	X	X					
Survey on Housing		X	X	X	X		X		X							
Survey on Parental Leave		X	X		X				X							
Survey on Transport		X	X	X	X		X		X							
Garda National Immigration Bureau	X	X	X	X		X	X	X	X		X				X	
Courts Tracking System		X	X				X									
Blind Person	X		X	X	X	X	X	X	X	X	X					X
Dependant Relative	X		X	X	X	X	X	X	X		X					X
Donations to Schools	X		X	X	X	X	X	X			X					X
Gift/Inheritance Tax Self-Assessment Return (IT38)	X		X	X	X	X	X				X					X
Health Expenses	X		X	X	X	X	X	X			X					X
Home Loan Interest	X		X	X	X	X	X	X			X					X
Homecarer credit	X		X	X	X	X	X	X	X		X					X
Housekeeper/incapacitated taxpayer	X		X	X	X	X	X	X	X		X					X
Incapacitated Child	X		X	X	X	X	X	X	X		X					X
Income Distribution of individuals on tax records	X		X	X	X	X	X	X			X					X
Inland Revenue Affidavit (CA24)	X	X	X	X	X	X	X				X					
Maintenance Payments	X		X	X	X	X	X	X			X					X
Medical Insurance	X		X	X	X	X	X	X			X					X
Particulars Delivered (PD) Form - ST21	X															
Permanent Health Benefit	X		X	X	X	X	X	X			X					X
Quarterly report of inflows and outflows to and from PAYE record	X				X	X	X									
Questionnaire Form CA4			X	X	X	X	X				X					
Rent Relief - Aged Persons	X	X	X	X	X	X	X	X	X		X					X
Rent Relief (non-aged persons)	X		X	X	X	X	X	X			X					X
Service Charges	X		X	X	X	X	X	X			X					X
Tax Population	X		X	X	X	X	X									X
Third Level Education Fees	X		X	X	X	X	X	X			X					X
Back to Education Allowance	X	X	X	X	X	X	X	X	X		X					
Back to Work Allowance	X	X	X	X	X	X	X	X	X		X					
Back to Work Enterprise Allowance	X	X	X	X	X	X	X	X	X		X					
Basic Supplementary Welfare Allowance	X	X	X	X	X	X	X	X	X		X					
Blind Person's Pension	X	X	X	X		X	X	X	X	X	X					
All Other Supplementary Welfare Allowance	X	X	X	X	X	X	X	X	X		X					
Carer's Allowance	X	X	X	X		X	X	X	X	X	X					

Data source title	PPS	Age	Gender	M/S[33]	S/E[33]	Income	G/C[33]	P/N[33]	F/S[33]	Dis-ability	Nation-ality	R/E[33]	T/C[33]	S/O[33]	R/A[33]	Farm
Central Records System (CRS)	X	X	X	X	X	X	X	X	X	X	X					
Child Benefit	X	X	X	X			X	X	X		X					
Disability Allowance	X	X	X	X		X	X	X	X	X	X					
Disability Benefit	X	X	X	X	X	X	X	X	X	X	X					
Exceptional Needs Payment, Back to School Clothing and Footwear Scheme, Urgent Needs Payment	X	X	X	X	X	X	X	X	X		X					
Farm Assist	X	X	X	X	X	X	X	X	X		X					X
Household Benefits Package and Free Travel	X	X	X	X		X	X	X	X	X	X					
Invalidity Pension	X	X	X	X		X	X	X	X	X	X					
Maternity Benefit	X	X	X	X	X	X	X	X	X	X	X					
All Occupation Injuries Benefit apart from Injury benefit	X	X	X	X		X	X		X	X	X					
Occupation Injuries Benefit - Injury Benefit	X	X	X	X	X	X	X	X	X	X	X					
Old Age Contributory Pension	X	X	X	X		X	X	X	X	X	X					
Old Age Non-Contributory Pension	X	X	X	X		X	X	X	X	X	X					
One Parent Family	X	X	X	X		X	X	X	X	X	X					
Pre-Retirement Allowance	X	X	X	X	X	X	X	X	X		X					
Retirement Pension	X	X	X	X		X	X	X	X	X	X					
SWA Rent Supplement	X	X	X	X	X	X	X	X	X		X					
Unemployment Assistance	X	X	X	X	X	X	X	X	X		X					
Unemployment Benefit	X	X	X	X	X	X	X	X	X		X					
Widows' and Widowers Contributory Pension	X	X	X	X		X	X	X	X	X	X					
Widow's Non-Contributory Pension	X	X	X	X		X	X	X	X	X	X					

Appendix 4 Data sources questionnaire

Each department was asked to complete this questionnaire for each of their key social data sources.

A Current situation	Notes
Title of data source	Title of administrative scheme or statistical survey.
Dept/Section/Agency responsible for data source	Name of the Section, agency, etc.
Contact details for person responsible for data source	Name, phone number, e-mail address.
Purpose of scheme	Brief summary of the main purposes.
Year of introduction	When scheme, survey, etc. was started.
Survey unit	The basic survey unit e.g. individual private household, farmer, specific individual, scheme applicant, hospital, etc.
Target population	This may be an aggregation of the survey unit or an aggregation of entities covered by the survey such as the number of eligible ewes for payment, total persons in households! Mention any controls such as upper or lower thresholds on eligibility using criteria su+B21ch as income etc. The target population includes eligible persons who do not apply to the scheme.
Sample size (Number of units surveyed)	This is the number of survey units included in the sample or scheme.
Response rate (if survey)	This will generally be units surveyed as a percentage of the target population but there may be complications around actual and usable response or eligible people who did not apply.
Data collection format	Paper questionnaire, online data entry etc. For administrative schemes, this generally relates to information completed by the applicant.
Is the dataset linked with an existing register	An example here would be DSFA scheme files may be linked to a central client file of all PPS Numbers. Scheme files may be held as separate annual files etc. but the register would cover all years.
How is data currently used - what statistics are produced - format of outputs	Refer to any statistical reports that are published or circulated.
Who are the users of the data	Make a general description here of the types of users plus identify any key users. Mention if some users require special analyses.
Is the dataset currently linked with others in/outside the Dept?	Indicate which other data sources it is linked with and what the relationship is. The link should be something fairly specific such as the link between certain DSFA schemes and the CSO Live Register.
The structure - computerised, software/system, how accessible, file format, etc.	If a scheme is operational in different local offices, mention if there is consistency across these local offices in how the information is computerised. 'File format' refers to the actual record layout. The intention here centres around whether the information could be easily exported into another system.
Was background/explanatory documentation on the data source received? Were coding manuals received?	Reviews of schemes, application guidelines, interviewer manuals, etc. 'Coding' refers to how the data is captured electronically. If the address coded to county level, which codes are used. Is occupation coded, etc.?
EU/national regulations or requirements	Such regulations generally indicate that there may be comparable data sources in other countries thus possibly increasing the importance of the source.

B Classification variables	The variables on the dataset - how they are collected, coded, are they stored electronically? Mention below if any of these classification variables are collected and not computerised. Mention also if the variable is badly completed or if it is often left blank. *[NOTE: Coding manual should be supplied to CSO]*
Title of data source	Title of administrative scheme or statistical survey.
Use of PPS Number	State whether the PPS Number is collected, what the completion/quality is like and whether a formal legal agreement exists to allow use of the PPS Number.
Use of other unique identifier	The identifier should be common across either files for different periods of the same scheme (longitudinal) or across schemes. An example is the Department of Agriculture herd number. Describe who generates the reference number and who the various users of it are.
Age/DOB	Describe the format age is held in and when it is captured. For example, is it coded as date of birth, or the age a person was when they applied to a scheme or an age-group range or that they qualify on age grounds for a scheme (over 65). Date of birth is the preferred format as it allows age to be generated on an ongoing basis and can assist in matching individuals.
Gender	Describe the format gender is held in and the coding system. For example, female=F and male=M. Mention if gender is not coded but can be deduced from the name of the person or the nature of the scheme.
Marital Status	Describe whether this information is collected and what answer options are allowed.
Socio-economic status	Mention if socio-economic status or group is collected and whether there are guidelines defining the individual categories e.g. skilled manual workers versus non-skilled manual workers.
Income	Describe the question asked, how income is defined, what period any income figures relate to (per week, the year prior to application to a scheme, etc.). Is this information updated on regular basis. Is it in punts or euro?
Geography	Is geography hard-coded or only derivable indirectly through the text information in a person's address. If coded, to what level (county, townland, urban/rural, etc.). Mention how these areas are defined e.g. urban versus rural and give copies of the coding systems to the teams. Mention if the information captured refers to somewhere other than main residence e.g. workplace or temporary/weekday residence. Mention also if the recorded/coded address represents the postal address or the grid reference address e.g. some households use a postal address of a neighbouring county.
Phone Number	If this is recorded, describe the format e.g. is area code asked for?
Family Status	Family status refers to whether a person has dependent children or has primary responsibility for an adult with a disability, who requires a high degree of support and attention.
Disability	Are disabled persons identified, i.e. persons having a total or partial absence of bodily or mental facilities, chronic disease, learning and personality disorders.
Nationality	Based on either country of birth or citizenship. Mention if the nationality is described e.g. Ireland or coded.
Race/Ethnicity	Includes race, colour and ethnic origin. Indicate which, if any, of these categories are recorded and how they are defined.
Membership of Traveller Community	Mention if travellers are identified, include those living in permanent accommodation.

Sexual Orientation	Mention if information is collected on sexual orientation, i.e. whether heterosexual, homosexual or bisexual.
Religious Affiliation	Is information collected on the religious belief to which a person adheres or the religious group to which he/she belongs?
Farm/Non-farm households	A farm can be defined in a number of ways. A farm is defined in the EU surveys on the structure of agricultural holdings as: a single unit, both technically and economically, which has a single management and which produces agricultural products. This definition is not based on income so a person owning a farm but earning the majority of his/her income from non-farming sources, would still be considered a farm as he/she is producing agricultural products. This distinction (production or main income source) has a major effect on the number of farms.
Other relevant variables	

C Quality Assessment	
Title of data source	Title of administrative scheme or statistical survey.
Completeness of records	Are the intended subjects of the survey or scheme covered? Is all the information specified on the initial recording document recorded? Is all the information captured initially entered on the computer record? This information can only be given by someone very familiar with the detailed processing of the data source file or in a position to run computer edits on the file. In practice, the section managing the data source will have to supply this information.
Duplicates	Do duplicate records exist? How are they removed?
Missing values	Are some items of information on the records left blank? If Yes, Why?
Other comments	
Duration of the data set (is it likely to continue?)	Is a long-term scheme or survey involved or ad-hoc once-off or periodic surveys/schemes? This information will be used as an input to assessing the potential value of the data source. One-off schemes, such as flood relief, have little statistical value.
Frequency of data - quarterly, annual?	Mention if differences occur in the level of coverage. For example, the mid-decade Census of Population is a less detailed questionnaire than the census undertaken at the start of each decade.
How regularly are new records added	New records relate to new scheme applications or additions to a survey population register. This may occur on an ongoing basis or at set periods.
How are inactive records treated	Inactive or dead records can be a particular problem. For longitudinal studies, it is desirable to keep these records but there should be a system of marking them as no longer within the scheme.
How are existing records updated	This refers to the up to dateness of the information. Some schemes may capture information at the time of first application and may not subsequently update this to reflect changes in age, educational qualifications etc.
How timely is the information?	This relates both to the up to dateness of the basic computer file and the timeliness of statistical reports.
Reliability of the reported data?	This can vary widely for different items of information e.g. age and income. Mention if computer edit or consistency check programs are run on the data source and if erroneous information is corrected or deleted.
Classifications used	Please list any international standard classifications used to code the data e.g. NACE Rev. 1.
Can results be extrapolated? Must the analyses be confined to the records included or can 'population' estimates be derived?	This question is related to earlier questions concerning samples and populations. If analyses are limited to survey returns and these are not representative then the analytical reports could be misleading.

D Future potential	The items in D are FOR COMPLETION BY CSO
Data source code	This code is generated by CSO and will be used to link different excel spreadsheets.
Title of data source	Title of administrative scheme or statistical survey.
Can useful additional information be produced from the existing data for internal or external use?	This item refers to whether more complex analyses could be undertaken than those currently published. These analyses may require more complex programming or involve merging data sources relating to different years to undertake a longitudinal analysis.
Are useful linkages possible between the existing data sets and other data holdings either internal or external?	This partially relates to cross-cutting issues. For example, would it be useful to link the data source with one from another department such as educational qualifications or to link data sources on a thematic basis across and within departments such as those related to disability.
How can the data sets be adapted to enhance their usefulness for statistics e.g. standard classifications, additional variables, unique identifiers to enable cross-linkages, etc ? There is need for clear demonstration of what results can be achieved and who will use them?	The use of standard classifications has a number of advantages: a centralised body such as the CSO may undertake all the work involved in compiling the classification/coding system; output tables can be produced using categories that are consistent with other reports e.g. age groups. Mention any data sources or variables that you know would add value to the data source e.g. nationality of parents of new babies.
Is there a potential use for enhancing CSO statistical compilation	This item relates to the value of the data source to existing CSO surveys. For example, information that would improve the compilation of the National Accounts or Balance of Payments.

Appendix 5 Age coding in selected data sources

Coding			Data sources
1	Date of birth		51
2	Age in years		2
3	Year of birth		2
4	Year and month of birth		3
5	Age bands	Child / Adult.	2
6		< 14,14 - 16,17 – 20, 21+.	1
7		15 - 24, 25 - 34, 35 - 49, 50 - 64, 64+.	2
8		< 40, 40 - 64, 65 - 69, 70 - 74, 75 - 79, 80 - 84, 85+.	1
9		Over 65.	1
	Total		**65**

Appendix 6 Gender coding in selected data sources

Coding		Data sources
2 categories (text/tick box)	Boy, Girl;	1
	Male, Female.	2
2 categories (numeric code)	Male = 1, Female = 2;	2
	Male = 2, Female = 1.	1
2 categories (letter code)	Male = M, Female = F;	14
	Male = M, Female = W.	15
3 categories (text/tick box)	Male, Female, Unknown.	2
3 categories (numeric code)	Male = 1, Female = 2, Unknown = 3;	2
	Male = 1, Female = 2, Unknown = 9;[34]	0
	Male = 1, Female = 2, Indeterminate= 3[35].	1
3 categories (letter code)	Male = M, Female = F, Unknown = Z;	1
	Male = M, Female = F, Unknown = NK;	1
	Male = M, Female = F, Unknown/Undetermined = U.	3
4 categories (text)	Male, Female, Indeterminate, Unknown.	1
4 categories (numeric code)	Male = 1, Female = 2, Indeterminate = 3, Unknown = 4.	1
Gender derived from forename		4
Gender derived from title		1
Revenue 'Personal Status' codes	Single Male case = 'A'; Single Female = 'B'; Married, wife earning = 'C'; Married, wife not earning = 'D'; Widower = 'E'; Widow = 'F'.	17
Total		**69**

[34] CSO Standard

[35] The category 'Indeterminate' refers to births where the sex of the child could not be determined. There were 10 such cases reported in live births in the Perinatal Statistics Report, 1993 by the Department of Health and Children.

Appendix 7 Marital status coding in selected data sources

Coding			Data sources
1	DSFA *ISTS* Coding	S = Single, M = Married, W = Widowed, D = Deserted, P = Divorced, C = Common Law / Cohabiting, Z = Separated, V = Unknown, L = Legally Separated.	25
2	DSFA *ISTS* Coding extended	S = Single, M = Married, W = Widowed, D = Deserted, P = Divorced, C = Common Law/ Cohabiting, Z = Separated, V = Unknown, L = Legally Separated, N = Never Married, T = Spinster, B = Bachelor, U = Unmarried.	1
3	Other letter coding	M = Married, D = Deserted, S = Single, W = Widowed, P = Separated, R = Religious.	1
4		S = Single, M = Married, P = Separated D = Divorced, W = Widowed, O = Other, U = Unknown/Undisclosed.	1
5		M = Married, S = Single, W = Widowed D = Divorced, E = Separated, O = Other, Z = Unknown.	1
6	Number coding	1 = Single, 2 = Married, 3 = Widowed, 4 = Separated / Divorced, 5 = Cohabiting.	1
7		1 = Single, 2 = Married, 3 = Widowed, 4 = Other (incl. Separated), 5 = Unknown, 6 = Divorced.	1
8		1 = Single, 2 = Married, 3 = Widowed, 4 = Divorced, 5 = Other, 6 = Unknown.	1
9		1 = Married, 2 = Never Married, 3 = Widowed, 4 = Separated, 5 = Divorced.	1
10		1 = Married, 2 = Never Married, 3 = Widowed, 4 = Separated, 5 = Divorced, 6 = Unknown; 7 = Single.	1
11		1 = Single, 2 = Married, 3 = Living as Married, 4 = Other, 9 = Don't know.	1
12		1 = Single, 2 = Married, 3 = Widowed, 4 = Common Law, 5 = Legally Separated, 6 = Other Separated, 7 = Deserted, 8 = Divorced, 99 = Other.	1
13	Text	Single, Married, Widowed, Divorced, Separated.	1
14		Single, Married (Incl. Common Law), Widowed, Divorced, Separated.	1
15		Single/Never Married, Married, Cohabiting, Widowed, Divorced, Separated.	1
16		Single; Separated; Divorced.	2
17		Cohabiting, Common Law, Deserted, Spouse, Divorced, Lone Parent, Married - Not Separated, Married - Separated, Single - Never Married, Unknown, Widow, Widower.	1
18	Title	Title	2
19	Revenue 'Personal Status' codes	Single Male case = 'A'; Single Female = 'B'; Married wife earning = 'C'; Married wife unearning = 'D'; Widower = 'E'; Widow = 'F'.	17
20	CSO Standard	1 = Single - never married, 2 = Married, 3 = Widowed, 4 = Divorced, 9 = Unknown. Codes 1, 3, 4 and 9 are not further disaggregated while code 2 may be diaggregated as follows: 21 = Married - not separated, 211 = Married - first marriage, 212 = Remarried - following widowhood, 213 = Remarried - following divorce/annulment, 22 = Married - separated, 2211 = Married - separated formally, 2212 = Married - separated informally.	0
	Total		**61**

Appendix 8 Nationality coding in selected data sources

Coding		Data sources
1	ISO 3166 Flat classification consisting of 239 categories. Countries are assigned a two letter code.	2
2	CSO standard Classification of countries based on the ISO-3166-1:1997. It is essentially a flat classification, but it includes a hierarchical element, which accommodates the constituent parts of the United Kingdom. There are 249 categories in total in this classification. These categories consist of the 239 ISO-3166 categories plus 10 additional categories broken down as follows: ▪ Two residual categories; ▪ Three categories for Guernsey, Jersey and the Isle of Man; ▪ Five sub-categories for the constituent parts of the United Kingdom.	0
3	Census of Population CSO standard classification of countries + 11 additional nationality categories.	1
4	GRO Flat classification with 242 categories. Countries are listed in alphabetical order and assigned a number from 1-244 (101 and 102 are excluded). The list of countries excludes four countries listed in ISO-3166, and includes five additional categories for the Channel Islands and the constituent parts of the United Kingdom. There is one residual category. This classification also contains two separate codes for Ireland and Éire.	1
5	ESRI codes – 10 categories, Ireland, Britain, and continents.	1
6	DES country codes - 3 digit numeric codes.	1
7	DSFA ISTS/CRS 38 categories – 37 country codes and one residual category.	28
8	Irish / British / Other EU / Non-EU.	1
9	National / Non-national.	2
	Total	**37**

Appendix 9 CSO Classifications and Standards Section

The Central Statistics Office has a statutory role to play in the co-ordination of official statistics in Ireland. In particular, it has the authority to ensure that appropriate standard classifications are used for this purpose. To this end, the Statistics Act 1993, 10(2), states that 'The Office (viz. Central Statistics Office) shall have authority to co-ordinate official statistics compiled by public authorities to ensure, in particular, adherence to statistical standards and the use of appropriate classifications.'

A classification is a useful tool for anyone developing statistical surveys. It is a framework which both simplifies the topic being studied and makes it easy to categorise all data or responses received. A classification consists of codes and descriptors and allows survey responses to be put into meaningful categories in order to produce useful data.

It is generally accepted that a standard classification will usually meet a number of main requirements that are outlined below.

Requirement	Description
Mutually exclusive categories	Each survey response should fit into only one category in the classification.
Exhaustive categories	All survey responses need to fit into the classification structure somewhere.
Precise and meaningful descriptors for categories	The content of each category in the classification should be clearly defined.
Conceptually sound	The classification should have a conceptual basis and a logical structure.
Statistically balanced	In general, survey responses should not fall heavily into one category and sparsely into the other categories.
Operationally feasible	There is no point in having a classification that cannot be implemented in practice.
Statistically robust	The classification should be able to be used for a number of years without revision.
Internationally comparable	The classification should be comparable with any international standard classification.

While many international standards exist, it is often necessary for individual National Statistical Institutes to adapt or modify these in order to cater for the needs of their own country or region. In Ireland, within the Central Statistics Office, a standard classification is one that has been approved by the Classifications Board as the official classification to be used for collecting, processing and outputting data on a particular topic. These standards are developed following extensive research of the topic in close consultation with stakeholders.

Benefits of Standard Classifications

1. Data integration: The use of statistical standards allows the integration of data from a variety of sources and between different periods of time.

2. Improved survey efficiency: Using standard classifications reduces the resources and time requirements associated with many aspects of survey development and maintenance. They assist in planning new survey developments and in reviews of existing surveys. Once a standard exists, it can be used in new surveys, thus eliminating the need to create afresh the conceptual definitions, code-files and coding rules that comprise the standard.

3. Meeting user requirements: Standard classifications are created in close consultation with stakeholders who are involved in developing statistical surveys. This ensures that classifications used to collect data in surveys will meet users' needs.

4. Data comparability: Standard classifications ensure that data is comparable over time and between different surveys, as they can be used to provide a link between different datasets.

5. Efficient use of resources: The use of standard code structures simplifies data processing and means only one version of the code numbers needs to be stored. Greater use of standard classifications increases and thus enriches the data produced by the Office.

6. More reliable statistics: The use of standard classifications gives reason for increased confidence in the reliability of statistics.

CSO standard classifications

The CSO Standard Sex Classification is a classification of sex based on the biological difference between male and female. This classification was approved by the CSO Classifications Board on June 12th, 2002. It is used within the CSO by the following sections: Census of Population, Household Budget Survey, Quarterly Industry, Quarterly National Household Survey, Tourism and Travel Section and Vital Statistics Section.

The CSO Standard Marital Status Classification is a four-level classification of marital status based on the marriage laws or customs of Ireland. This classification was approved by the CSO Classifications Board on June 12th, 2002. This classification will be used by the Quarterly National Household Survey (QNHS), Household Budget Survey (HBS), and Census of Population Surveys. It has links with the EU SILC Marital Status Classification and the General Registration Office (GRO) Marital Status Classification.

The CSO Standard Country Classification is a classification of countries based on the ISO-3166-1 (1997). The ISO 3166-1 is an international classification developed by the International Standards Organisation (ISO), and based on a listing of country categories maintained by the United Nations Statistics Division. It must be reviewed at no longer than 5-year intervals and updates are made available on the Internet. It is essentially a flat classification, but it includes a hierarchical element to accommodate the constituent parts of the United Kingdom. There are 249 categories in the classification in total. This classification was approved by the CSO Classifications Board on May 6th, 2003 and will be adopted by the following areas in the CSO: Annual Services Inquiry, Balance Of Payments, Business Register, Census of Industrial Production, Census Of Population, Data Office, EU-SILC, Quarterly National Household Survey, Tourism, Transport and Trade.

The CSO Standard Classification of Industrial Activity, NACE Rev.1.1, is a statistical classification of economic activities (General Industrial Classification of Economic Activities within the European Communities). NACE Rev.1.1 is designed to allow comparisons of economic activities at national, European and global levels. On January 1st, 2003 NACE Rev.1.1 became an official CSO standard classification. The following is a list of the classifications with official links to NACE Rev.1.1:

- ISIC Rev.3.1 (International Standard Industrial Classification of All Economic Activities, Third Revision);
- CPA (Statistical Classification of Products by Activity in the European Economic Community);
- NACE 70 (General Industrial Classification of Economic Activities within the European Communities, NACE 70);
- CPC (Central Product Classification);
- HS (Harmonised System);
- CN (Combined Nomenclature); and
- PRODCOM (The List of Products of the European Community).

Appendix 10 UN fundamental principles of official statistics

The Economic Commission for Europe[36]

Bearing in mind that official statistical information is an essential basis for development in the economic, demographic, social and environmental fields and for mutual knowledge and trade among the States and peoples of the world.

Bearing in mind that the essential trust of the public in official statistical information depends to a large extent on respect for the fundamental values and principles which are the basis of any society which seeks to understand itself and to respect the rights of its members.

Bearing in mind that the quality of official statistics, and thus the quality of the information available to the Government, the economy and the public depends largely on the co-operation of citizens, enterprises, and other respondents in providing appropriate data needed for necessary statistical compilations.

Recalling the general provisions and standards adopted to this end by the European Convention on Human Rights, the Convention of the Council of Europe of 28 January 1991 for the Protection of Individuals with regard to automatic processing of personal data, the Final Act of the Helsinki Conference on Security and Co-Operation in Europe and the Charter of Paris for a New Europe,

Recalling the efforts of governmental and non-governmental organisations active in statistics to establish standards and concepts to allow comparisons among countries:

Recalling also the International Statistical Institute Declaration of Professional Ethics,

Having taken cognisance of the consensus reached within the Conference of European Statisticians on the need to define the principles governing the activities of the official statistical agencies in the region and in the member States,

Adopts the present resolution:

Principle 1: Official statistics provide an indispensable element in the information system of a society, serving the government, the economy and the public with data about the economic, demographic, social and environmental situation. To this end, official statistics that meet the test of practical utility are to be compiled and made available on an impartial basis by official statistical agencies to honour citizens' entitlement to public information.

Principle 2: To retain trust in official statistics, the statistical agencies need to decide according to strictly professional considerations, including scientific principles and professional ethics, on the methods and procedures for the collection, processing, storage and presentation of statistical data.

Principle 3: To facilitate a correct interpretation of the data, the statistical agencies are to present information according to scientific standards on the sources, methods and procedures of the statistics.

Principle 4: The statistical agencies are entitled to comment on erroneous interpretation and misuse of statistics.

[36] Decision C(47) adopted by the Economic Commission for Europe at its April 1992 session.

Principle 5: Data for statistical purposes may be drawn from all types of sources, be they statistical surveys or administrative records. Statistical agencies are to choose the source with regard to quality, timeliness, costs and the burden on respondents.

Principle 6: Individual data collected by statistical agencies for statistical compilation, whether they refer to natural or legal persons, are to be strictly confidential and used exclusively for statistical purposes.

Principle 7: The laws, regulations and measures under which the statistical systems operate are to be made public.

Principle 8: Co-ordination among statistical agencies within countries is essential to achieve consistency and efficiency in the statistical system.

Principle 9: The use by statistical agencies in each country of international concepts, classifications and methods promotes the consistency and efficiency of statistical systems at all official levels.

Principle 10: Bilateral and multilateral co-operation in statistics contributes to the improvement of systems of official statistics in all countries.

The conference of European Statisticians, at intervals of not more than three years, will discuss these principles, consider ways to contribute to their application and report to the Commission.