
Chair of the UK Statistics Authority, Sir Michael Scholar KCB

Rt. Hon. Francis Maude MP
Minister for the Cabinet Office
Cabinet Office
70 Whitehall
LONDON
SW1A 2AS

16 March 2012

Dear Francis

STATISTICS FROM ADMINISTRATIVE DATA

I am enclosing a paper by the Statistics Authority that looks at some of the implications of the growing use of administrative records to prepare official statistics.

This is a think-piece. I believe we need it, because we need to take stock of where we are, and where we're going on this important matter.

The trend towards a growing statistical use of administrative data is very much to be welcomed. It saves money compared with conventional surveys, and the statistics are often more complete and available for smaller geographical areas. They can also be joined up with records from other sources to give a richer basis for analysis. And this trend serves the Open Data agenda by ensuring that confidential personal and business records are turned into publicly usable statistical material in a proper professional way.

At the same time, we have to recognise that there are some vulnerabilities inherent in the statistical use of administrative records and that there is a need for strong safeguards and guarantees about the protection of personal and confidential information.

As the paper says, the future value of the statistical service to government and the public will depend, in large part, on how effectively we are able to exploit existing and future administrative records for statistical purposes. There are many reasons why this is true. It is partly that we should make best use of what the government already holds to keep costs down; partly that we should exploit the opportunities presented by web based provision of public services; and it is partly about the growing costs and challenges of conventional surveys.

We offer some conclusions on the way forward. In essence:

- the government needs to find ways to make it easier for statisticians to identify and use administrative sources that have the potential to produce valuable statistics;
- bodies that control those administrative sources need to reduce the barriers to statistical use;
- arrangements for cost-effective statistical audit (checking that records are what they appear to be) need to be strengthened;
- definitions and standards in different administrative sources need to be harmonised where possible;
- the nature of the confidentiality safeguards for sharing administrative data for statistical purposes should be more openly debated; and,

- where administrative systems are used to produce statistics, statisticians should be centrally involved in the design and operation of those systems. They would add value, help to secure the quality of the data, and ensure that statistical producers are aware of data changes.

This agenda is all the more important in the context of the work ONS is doing to find a substitute for the traditional decennial Census that relies to a much greater extent on administrative data. The Statistics Authority would be glad to be involved in further deliberations to achieve progress in these matters.

I am copying this letter and enclosure to Jil Matheson, Jeremy Heywood and Tim Kelsey.

Yours sincerely

A handwritten signature in black ink that reads "Michael Scholar". The script is cursive and fluid.

Sir Michael Scholar KCB

Monitoring Brief

Monitoring Brief 3/12
16 March 2012

Creating official statistics from administrative data

Contents

Section 1	Summary
Section 2	Findings
Annex 1	Evidence and analysis
Annex 2	The use of administrative sources for statistical purposes in the UK and internationally
Annex 3	Navigating the legal obstacles
Annex 4	Statements of Administrative Sources
Annex 5	The implications of Open Data policy

Summary

1. This Brief reviews whether the statistical service has sufficient access to, and influence over, the administrative data sources from which official statistics are increasingly drawn. It also looks at some of the statistical implications of the Government's Open Data policy. It concludes that in order to ensure quality and maximise value, some further steps need to be taken - both to improve access for statistical purposes and to build confidence that appropriate security arrangements exist to protect confidential information about individuals and businesses.
2. Virtually all of the activities of the state that involve the provision of services to the public or businesses now use computerised administrative systems. The resulting data - essentially the records of transactions - offer great potential for the creation of robust and detailed statistics. This has led to considerable growth in the range of official statistics available over recent decades and a marked shift in the balance between statistics based on survey data and those based on administrative data. There will always be a role for sample surveys but much of the work of government statisticians is now about managing and analysing data from administrative sources.
3. Administrative data are now the source of a very wide range of UK statistics on: demography; the economy; agriculture; public services; deprivation and health. They have a number of strengths, compared with statistics from sample surveys:
 - reduced burden on respondents;
 - low cost production processes;
 - better quality statistics for small geographic areas;
 - more frequent updating, usually; and
 - the capacity to link information from different sources to produce richer data.

But there are also some limitations:

- potential inconsistency between organisations that supply the data;
- inconsistency over time as the administrative rules change;
- incompleteness – the administrative system may not cover the whole of the population to which it relates;
- mismatches may occur between what the user of statistics requires and what can be produced from the most apposite administrative data; and
- it may be impossible to operate the sorts of controls on pre-release access that help to prevent misuse.

Notwithstanding the disadvantages cited above, the huge benefits will ensure that administrative data become central to the production of official statistics. The need therefore is to find ways to ameliorate the disadvantages.

4. In the light of the research outlined later in this report, our main conclusions are that:
 - a. A more systematic and thorough process is required for identifying administrative databases that have potential to be used to produce official statistics. The proposed Public Sector Information Inventory should be used to identify administrative sources that have the potential to be used for statistical purposes (paragraphs 9-10 and A1.1 – A1.4).
 - b. Ways need to be found to simplify the process of giving statisticians access to administrative data held in organisations other than the one that the statistician happens to work in. Public bodies that control administrative data should be required to publish plans to remove barriers to data sharing for statistical purposes (paragraphs 11-16 and A1.5 – A1.12).
 - c. Administrative data are simply a by-product of administrative processes and may not exactly match the needs of the statistician or the user of statistics. There is a need for more central guidance to bodies that produce statistics from administrative data about auditing and ensuring the quality of those data (paragraphs 17-20 and A1.13 – A1.20).
 - d. The utility of administrative data as sources of statistics can be limited by, for example, inconsistencies in the data held by organisations responsible for similar types of administration, and by changes over time that result as, for example, definitions or eligibility rules change. There is a need to review the implications for official statistics of the current use of different standards, definitions and classifications in administrative systems (paragraphs 21-24 and A1.21 – A1.23).
 - e. Some people believe that data sharing runs counter to civil liberties and personal privacy. There would be value in a public debate, perhaps in the context of ONS' *Beyond 2011 Programme*, led by an authoritative and independent figure, about the appropriate safeguards for the sharing of administrative data for statistical purposes (paragraph 25).

Findings

5. The future value of the statistical service to government and the public will depend, in large part, on how effectively and efficiently it is able to exploit the administrative data of government for statistical purposes. To maximise this use will require the strongest possible guarantees for the individual person and business that their confidential information will not be made public or misused in the course of creating aggregate statistics from the original records. The balancing of, on the one hand, the need to provide administrative records to statistical offices with, on the other, the need to ensure complete confidentiality, presents real challenges.
6. The answer is not to lock up the administrative records in the organisation which first produced them so tightly that they cannot subsequently be used for the creation of statistics. Rather, it is a matter of ensuring that statistical offices have all the right safeguards in place, and steadily building public confidence in their ability to handle confidential records. In practice, the Office for National Statistics and other statistical offices in government already routinely process tens of millions of such records, and do so professionally and successfully. As well as robust technical security measures and tight restrictions on who within these offices is allowed to inspect records, statistical offices have the advantage of not being interested in the *individuals* whose records they hold. Their work is about producing statistical aggregates of different kinds, and not about using information from individual records. The security arrangements can be designed to reflect that and be the more robust because of it.
7. Over many years, concerns about the security of administrative data have led to a mass of UK statutory and other restrictions on access to them. Often these restrictions have been put in place without regard to the need to provide access for ONS and other statistical offices. In some cases the restrictions are so tight that government statisticians do not know, in any detail, what the administrative records contain. This makes it difficult to evaluate them for statistical use, thus creating further obstacles. We note (paragraph 19) that European legislation to facilitate the use of administrative data for statistical purposes is under discussion.
8. The increasing demand for statistics that will support evidence-based decision-making at detailed geographic levels, at a time of tight public expenditure control, has created a strong case to maximise the statistical value of administrative data. Faced with the statutory and cultural barriers that exist, we support the creation of the UK Administrative Data Task Force which we understand will look at barriers to data sharing during 2012.

Identifying suitable administrative sources

9. A more systematic and thorough process is required for identifying administrative databases that have potential to be used to produce official statistics. The *Code of Practice for Official Statistics*¹ requires bodies that produce statistics to “evaluate existing data sources ... before undertaking new surveys”². There therefore needs to be a way to locate and evaluate the statistical potential of all the administrative databases that exist across government.
10. Government statisticians will normally be aware of the main administrative sources which exist within the organisations in which they themselves work; but they may not know about relevant data held by other parts of government. In the light of the Government’s Open Data policy, the existence and a description of the content of administrative databases managed by public bodies should be public information and readily found. However, this does not yet seem to be the case – we note that similar exercises have been attempted in the past - yet,

¹ <http://www.statisticsauthority.gov.uk/assessment/code-of-practice/index.html>

² Principle 6, Practice 5 of the *Code of Practice*

until such time as this is addressed successfully, the statistical service will need to find its own ways to identify and share information about administrative sources across government.

Accessing administrative sources for statistical purposes

11. Ways need to be found to simplify the process of giving statisticians access to administrative data held in organisations other than the one that the statistician happens to work in.
12. Where statisticians work within the organisation that controls an administrative database, their access to those data for statistical purposes is a matter of seeking the necessary approval from the senior management of the organisation.
13. But where the administrative database is controlled by another organisation, data access is more problematic. Whilst there is rarely a problem with getting hold of *aggregate* statistics from other departments and agencies, access to the detailed records (for statistical purposes) can run up against statutory and administrative obstacles related to the protection of personal and business confidentiality.
14. Although the *Statistics and Registration Service Act 2007*³ enables unit-level data to be shared between a Crown body and the Office for National Statistics, it must be for a specified purpose and be subject to Parliamentary approval. Moreover, the provisions of the Act only enable such sharing; they do not oblige the holder of the data to share it. This means that the body that controls the data can refuse to share it, even if Parliamentary approval is granted.
15. Assessing the statistical potential of an administrative source and gaining access permissions from both Parliament and the data custodian can therefore be a lengthy and costly process; and it may not be possible to assess the potential value of the data for statistical purposes until *after* access has been granted. That can impede the preparation of the business case. This is a Catch 22 model of data sharing. Until the statisticians are given access to an administrative database they may not have the evidence they need to make the case to be given access.
16. Despite the existence of useful guidance on the legal framework for data sharing issued by the Ministry of Justice⁴, many departments have difficulty in navigating the legal obstacles. These difficulties are discussed further in Annex 3.

Assuring quality in relation to statistical purposes

17. The Assessment reports of the Statistics Authority frequently raise questions about the consistency, completeness or quality of data from administrative sources on which National Statistics are based. Producer bodies should seek to offer a more systematic assessment of the quality of the underlying administrative data. In a sample survey, the questions asked can be tailored specifically to the subject of interest. For example, a housing survey questionnaire could ask about the number of bedrooms in a home. However administrative data are simply a by-product of administrative processes and may not exactly match the needs of the statistician or the user of statistics – the administrative data might give the address and value of the property allowing only a rough estimate to be made of the number of bedrooms. So there are several different aspects to ensuring quality:
 - What is recorded in the administrative records may differ from the concepts and definitions of statistical interest; and judgements need to be made as to how reliable the eventual statistics are going to be.

³ The Statistics Act governs data sharing between ONS and other public bodies.

⁴ <http://www.justice.gov.uk/guidance/freedom-and-rights/data-sharing.htm>

- The way in which the administrative systems are operated, or definitions interpreted, may affect the suitability of the data for producing statistics. So it is important to be able to check that all the administrative data (which might come from schools, hospitals or offices all over the country) are using the same definitions and procedures.
 - The statisticians also need to be sure that having identified any concerns about definitions or consistency in the administrative data system, they will have sufficient influence over the management of the system to ensure that matters that need to be addressed can be resolved. It is not uncommon for statisticians to know of problems with administrative systems they are using to produce statistics but not be in a position to ensure that they are resolved.
18. Assessment reports also suggest that many producers do not address the quality of administrative sources in the way required by the *Code*. The *Code* expects each producer of official statistics to publish a Statement of Administrative Sources⁵ - see annex 4 - but many of these Statements are not as well researched as they might be. In particular, there are shortcomings in relation to administrative sources that are controlled by a different organisation from the one that produces the statistics.
19. As part of its policy on Robust Quality Management⁶, Eurostat - the EU statistics office - will bring forward legislation to require Member States to make public 'Commitments of Confidence' in their European Statistics. The precise format of the statement is not yet known, but the intention is to require producers of European Statistics to establish the quality of "upstream" administrative data, and comment on the impact that this has on the published results.
20. It is central to the Government's Open Data policy that administrative databases should be published and used by the public to hold providers of public services to account. However, this means that the completeness, quality and consistency of those data will be critical. Some of the implications of the policy are considered further in Annex 5.

Influencing administrative data holdings to enhance their statistical potential

21. It was noted earlier that the utility of administrative data as sources of statistics can be limited by, for example, inconsistencies in the data held by organisations responsible for similar types of administration, and by changes over time that result as, for example, definitions or eligibility rules change. These are substantial issues which tend to require strong co-ordination and multi-agency support to resolve, if coherent and consistent statistics are to be produced. Whilst the Government Statistical Service has been able to make considerable progress in the harmonisation of survey questions and classification, corresponding approaches in relation to administrative data have made less progress – because of the departmental ownership of the administrative systems.
22. The *Beyond 2011 Programme*⁷ is currently exploring ways to meet UK users' needs for population and other Census statistics using (primarily) administrative data. Denmark was the first country in the world, in 1980, to conduct a 'census' using just administrative data⁸, followed in 1990 by Finland. The United Nations Statistics Division estimates that the 2001 Finnish 'census' cost just €0.2 per head, compared to an estimated €6.2 per head in the UK⁹,

⁵ Under Protocol 3, Practice 5e of the *Code of Practice*.

⁶ http://epp.eurostat.ec.europa.eu/portal/page/portal/quality/documents/COM-2011-211_Communication_Quality_Management_EN.pdf

⁷ <http://www.ons.gov.uk/ons/about-ons/what-we-do/programmes---projects/beyond-2011/index.html>

⁸ http://www.nordbotten.com/articles/Adm_data.pdf

⁹ [http://unstats.un.org/unsd/economic_stat/intl%20coop%20and%20workshops%20\(bes\)_files/Pretoria-Agenda_files/notes/UNSD_AdminData.ppt#12](http://unstats.un.org/unsd/economic_stat/intl%20coop%20and%20workshops%20(bes)_files/Pretoria-Agenda_files/notes/UNSD_AdminData.ppt#12)

where such techniques cannot be used because successive governments have been unwilling to introduce a Population Register.

23. The UK Statistics Authority endorsed the importance of administrative sources for statistical work in the *Code of Practice for Official Statistics*. The *Code* introduced the requirement for each producer of official statistics to prepare and publish a Statement of Administrative Sources, intended in part to stimulate positive change in the way that administrative data are identified and used (by themselves, in combination with other sources, and for modelling purposes) by producers of statistics.
24. There is evidence that some government departments did not at first see much value in the Statements of Administrative Sources – citing low numbers of website ‘hits’ as evidence. However, the evolving dialogue between the Authority and departments has helped to clarify the purpose and potential value of Statements, and we have seen some Statements that show a more complete consideration of quality issues and change control processes than was demonstrated previously. For example, the Information Centre for Health and Social Care has recently updated its Statement¹⁰ to include detailed information about all of the administrative sources that it uses. The Authority considers this to include all of the necessary information required for *Code*-compliance and commends the Information Centre.
25. Pressures on public finances and the increasing need to provide statistics that enable evidence-based decision-making at detailed geographic levels have placed a renewed emphasis on the need to maximise the statistical value of administrative data. The Authority recognises the legal, quasi-legal and ‘cultural’ barriers that exist to the realisation of the full statistical potential of administrative data, and considers that:
 - more needs to be done to overcome these barriers – we understand that a UK Administrative Data Task Force will look at both legal and ‘cultural’ barriers to data sharing during 2012, activity which we believe is very important to enhance the statistical potential of administrative data;
 - further thought should be given to stimulating an increase in the value of existing (and new) administrative data – we have been told that Scottish Government statisticians have been working with other organisations to help them obtain value from the data they hold: by organising secondments for Scottish Government statisticians, and holding workshops and seminars. These seem constructive and helpful activities. More generally, where administrative systems are used to produce statistics, statisticians should be centrally involved in the design and operation of those systems. They would add value, help to secure the quality of the data, and ensure that statistical producers are aware of data changes; and
 - the main underlying issue – the feeling amongst some of the public that data sharing runs counter to civil liberties and personal privacy – should be addressed head on, by stimulating an open debate. This could take the form of a consultation, perhaps in the context of *Beyond 2011*, that sets out the benefits of data sharing for statistical purposes, and that seeks to establish an informed view about whether the privacy safeguards that statisticians work within are consistent with the acceleration of data sharing that seems vital for the future development of official statistics. Given the nature of the subject, such a debate would most appropriately be led by an authoritative and independent figure.

¹⁰ <http://www.ic.nhs.uk/statistics-and-data-collections/publications-calendar/administrative-sources/list-of-administrative-sources>

Annex 1

Evidence and analysis

Identifying suitable administrative sources

A1.1 Producers are required¹¹ to identify sources of administrative data with statistical potential, and to document these within their Statement of Administrative Sources. However, our review of these Statements has suggested that producers are often unsuccessful in identifying sources of administrative data that might be appropriate for statistical use. In large part this seems to be because there is no centralised source of information in the UK which could help producers of official statistics to identify suitable administrative data. Whilst the Administrative Data Liaison Service (ADLS – see A2.4) provides an analogous function within academia, there is currently no equivalent organisation or group within the Government Statistical Service (GSS).

A1.2 Many other national statistical systems are supported by legislation that enables data sharing and are thus able to maximise the use and value of administrative data holdings. For example, the Statistics Act 1993¹² in Ireland gives CSO Ireland far-reaching powers to access administrative data for statistical purposes. The CSO examined¹³ the statistical potential of the data holdings of six government departments, following a directive that the CSO should take the lead in developing the potential of administrative data sources across Government. The CSO's review culminated in a set¹⁴ of recommendations intended to maximise the statistical potential of administrative data by encouraging a joined up statistical system characterised by:

- Collaboration between government departments to ensure that statistical potential is realised efficiently¹⁵; and
- CSO taking a lead role in harmonising classification and coding systems and promoting, maintaining and disseminating their use¹⁶.

A1.3 More recently, in the UK the Transparency Board has recommended that there should be a single Public Sector Information Inventory. This is not just about data available to the public – it would be an inventory of all public sector information. The inventory itself will be a public document. The public will be invited to challenge whether those items on the list that are closed to public use should in fact be opened.

A1.4 The Statistics Authority supports the systematic identification of data holdings, and the scrutiny of the proposed inventory with a view to supporting the further use and enhancing the value of administrative data for statistical purposes.

Accessing administrative sources for statistical purposes

Sharing data with ONS

A1.5 It can be difficult to evaluate administrative data held by other public bodies, sometimes because access to the data – even for the purposes of evaluation – is prevented. ONS has managed to overcome this difficulty in one instance by working in partnership with

¹¹ Under Protocol 3, Practice 5c of the *Code of Practice*

¹² <http://www.irishstatutebook.ie/1993/en/act/pub/0021/print.html#sec11>

¹³ <http://www.cso.ie/en/releasesandpublications/corporatepublications/statisticalpotentialofbusinessenvironmententerprisedataholdingsinselectedgovernmentdepartments-workingreport/>

¹⁴ The review proposed 12 general recommendations and a further 9 specific recommendations directed at individual government departments.

¹⁵ Recommendation 2.

¹⁶ Recommendation 5.

another statistical producer. ONS is in the process of acquiring some VAT data that are held by Her Majesty's Revenue & Customs (HMRC), with a view to improving the Annual Business Survey's (ABS) edit rules and to analysing the relationship between VAT and Turnover in the National Accounts. Since access to the data was initially restricted to HMRC employees, ONS found it difficult to determine whether the value of the data merited incurring the costs associated with gaining data sharing permissions under the *Statistics Act*. To overcome this, ONS arranged for an individual to be seconded to HMRC. The ONS employee was then able to access HMRC data in order to determine whether the data were useful enough for ONS' intended purposes. HMRC was aware and supportive of the data access element of the secondment.

- A1.6 Whilst we recognise the co-operation in this case, it can hardly be regarded as an optimal solution. It is unfortunate that there is no provision for pilot studies in the legal framework for data sharing. There may be scope however for greater use of ONS' Virtual Microdata Laboratory¹⁷ (VML) in this regard, for example, by establishing remote access to ONS' VML in every GSS department, or by departments establishing their own VML and allowing other departments to access them. And there may be scope for more use of common law powers, to enable limited scope one-off studies that might be regarded as 'proportionate'.
- A1.7 We are aware of a few examples of ONS having successfully arranged access to administrative data for statistical purposes, via Information Sharing Orders (secondary legislation, under the terms of the *Statistics Act*). For example, ONS has arranged access to the Department for Education's (DfE) individual level School Census Data to improve its migration estimates. This arrangement was developed in response to the 2006 Interdepartmental Task Force on Migration which proposed that "Access should be provided to individual level School Census data for statistical purposes, to allow linking with other sources so as to improve the statistics on migrants and their families..."¹⁸. ONS has also successfully acquired PAYE data from HMRC, for use as a sampling frame for the Annual Survey of Hours and Earnings (ASHE).
- A1.8 However, such instances of data sharing are infrequent – despite the fact that when personal administrative data are shared with ONS, they benefit from the non-disclosure rule (under section 39 of the *Statistics Act*) and the criminal offence of wrongful disclosure. For some (perhaps most) administrative data this is the first time that they are protected by a specific criminal sanction for wrongful disclosure. In those circumstances the protections are strengthened.

Sharing data across government

- A1.9 Data sharing mechanisms apply within the broader context of the GSS; government departments are part of the same statistical system, producing official statistics in line with shared principles and processes. When administrative data are used for statistical purposes, the Code of Practice applies. This brings a regulated regime for disclosure protection, including both protection in outputs, and protection from non-statistical requests. When administrative data are used for statistics and research the Data Protection Act requires that no measures or decisions with respect to a particular individual can be undertaken, and also that no damage or distress be caused. So in this

¹⁷ The Virtual Microdata Laboratory (VML) is a facility within the Office for National Statistics (ONS) which enables access to restricted microdata for research purposes. Researchers from government and academia use the VML to carry out research on ONS surveys and other confidential datasets. Using VML means that data are not transferred from one organisation to another: the raw data do not migrate from one system to another but the user is able to query the data remotely and obtain aggregate results printed out on his/her own system.

¹⁸ <http://www.ons.gov.uk/ons/guide-method/method-quality/imps/updates-and-reports/historical/2009/research-paper-on-the-use-of-school-census-data-to-improve-population-statistics---october-2009.pdf>

context, the GSS maintains a vast resource of administrative data, all of which are kept and used in line with confidentiality and security arrangements that stem from the same requirements and principles.

- A1.10 Business and public sentiment about data sharing seems to be mixed - many expect official bodies to share data in order to improve service delivery and minimise form-filling and bureaucracy, whilst others have reservations about data sharing based on considerations of privacy and civil liberties. Setting this debate to one side, Government is mindful of the need to protect the security of information about individuals; the Cabinet Office requires departments to assign their information assets an Impact Level¹⁹. But there is a potential risk that this regime may be used as a barrier to data sharing – for example, by one department declaring that its data has such a high Impact Level that its acquisition by another department, all other things being equal, would be prohibitively and disproportionately expensive. This is one example of an area in which Government might usefully take a central view of risks and benefits of data sharing for statistical purposes. The adoption of shared data centres, governed by a ‘safe haven’ concept of an agreed information security management standard, might remove unwarranted differences in impact level assessments. It might also make data sharing a matter of changing access permissions within the data centre, rather than the current and somewhat archaic practice of physically moving data from one place to another when it is shared.
- A1.11 It seems that the time and cost involved, and the need to negotiate the complex legislation that governs data sharing between producers within the UK statistical system, are justifiably seen as barriers. These barriers are not new. Overcoming them should be an obligation on a department when the benefits seem to outweigh the costs, subject to an analysis of the risks involved. It would be helpful if departments and agencies that hold administrative data were to publish plans to remove barriers to data sharing with bodies that produce official statistics.
- A1.12 Once these plans have been developed, and the full range of barriers and potential solutions are better understood, the Statistics Authority will be able to advise Government about whether it sees a case for the reform of the legal arrangements for data sharing across government²⁰.

Assuring quality in relation to statistical purposes

- A1.13 Producers are required to demonstrate that administrative data are subject to appropriate quality audit processes by documenting these processes within their Statements of Administrative Sources. Our review of producers’ Statements identified that, often, producers do not account for all dimensions²¹ of quality within their audit process. But adopting a holistic view of quality is important for two reasons:

¹⁹ For example, Impact Level 0 is “no impact, public access”; the highest Impact Level is IL6.

²⁰ For example, the Authority notes that section 43 of the *Statistics Act*, which opens a permanent gateway for information relating to NHS registration to be sent to ONS, might set a valuable precedent for sharing administrative data, for statistical purposes, across government:

- the gateway is specific to the data, not to who holds it. In contrast, an Information Sharing Order is specific to the supplier.
- the gateway covers the ‘index’ variables of address, previous addresses, age, sex, and patient identification number. It excludes any contextual information such as the patient’s record of health or social care.
- some of the exceptions to the non-disclosure rule are disabled.

²¹ The six dimensions of the ESS Quality Framework are: relevance, accuracy, timeliness and punctuality, accessibility and clarity, comparability and coherence.

- It allows producers to identify quality issues that might otherwise not be detected – for example differences and inconsistencies in the way that variables are defined or categorised; and
- It allows producers to make informed judgements about the quality of an administrative source and its suitability for a particular statistical use.

A1.14 The risk of an incomplete consideration of quality was highlighted in the Authority's assessment of the Scottish Government's statistics on children who are in contact with child protection services²². Inconsistencies between the practices in different local authorities meant that the resulting (aggregated) statistics were not suitable for making comparisons between local authorities, nor for producing consistent national totals.

A1.15 Quality issues such as these might not be detected unless producers ensure that the operation of the administrative processes, 'on the ground', are consistent with their expectations and assumptions in producing statistics that are based upon data resulting from these processes. We understand that ONS is closely involved, alongside some other European National Statistical Institutes, in a virtual network to find common ways to use administrative data for business statistics and to share the best practices and methods developed with all interested member states²³.

A1.16 One strand of the work is to review how the existing European Statistical System (ESS) quality framework can be applied when administrative data are used as inputs; the origins of the current quality framework lie in the use of survey data. So for example indicators relating to 'accuracy' need to be defined in another way (than sampling errors). The scope of the work is presently limited to quality indicators that are directly impacted by the use of administrative data; the ESS quality dimensions of accessibility and clarity and relevance are not covered.

A1.17 The issues underpinning the need for 'quality audits' are being addressed in part by this European project by developing checklists (by mid 2013) relating to:

- the issues that a statistical producer needs to think about when considering administrative data options or if there are changes to administrative data that they already use; and
- the quality of the administrative data when it arrives with the statistical producer. This is in terms of checking that the data include the information required, are in the appropriate format, relate to the appropriate time periods, and so on.

A1.18 Other activity, under a Blue-ETS²⁴ project which builds on the work of Statistics Netherlands, looks at the quality of administrative data in terms of sources, metadata and data; it is mainly concerned with considering which administrative data sources fulfil the requirements of the statistical producers. The aim is to produce a Quality Report Card for administrative data.

A1.19 We have been told that if these administrative data checks are implemented, these will give statistical producers some awareness of the issues related to 'quality audits' – they would ensure that some of these crucial questions are asked at the beginning of the process rather than only after the output is produced.

A1.20 Whilst we support this activity, we are concerned that an over-reliance on a checklist approach may lead some producer bodies to over-simplify the underlying issues, and to a situation in which the existence of different administrative processes is recognised but

²² These statistics were the subject of assessment report number 67. The report can be downloaded at: <http://www.statisticsauthority.gov.uk/assessment/assessment/assessment-reports/index.html>

²³ <http://essnet.admindata.eu/>

²⁴ <http://www.blue-ets.istat.it/index.php?id=40>

where there is no obvious responsibility for the remedial activity required to develop consistent and comparable statistics. (We do however recognise that such checks are a valuable stimulus to identifying the need for further action). To determine fully the quality of administrative data, producers need an in-depth knowledge of the source, including its primary (administrative) purpose and the way that the data are collected and processed; this might usefully be supported by central guidance to producers of official statistics about auditing the quality of administrative data for statistical purposes. This guidance could distinguish between statistical quality audits, on the one hand, and 'traditional' audits on the other. For example, a 'traditional' audit of a police force's recorded crime statistics might concentrate on whether the recorded figures matched the police activity; a statistical quality audit, on the other hand, would look across police forces to ensure that crimes and incidents were being recorded consistently.

Influencing administrative data holdings to enhance their statistical potential

A1.21 The Code of Practice addresses this issue from a number of different perspectives – it requires bodies that produce official statistics to:

- “promote statistical purposes actively in the design of administrative systems in order to enhance the statistical potential of administrative records” (Principle 6, practice 3)
- “maximise opportunities for the use of administrative data” (Protocol 3, practice 3); and
- “prepare ... a Statement of Administrative Sources ... which identifies the ... procedures to be followed within the organisation to ensure that full account is taken of the implications for official statistics when changes to administrative systems are contemplated” (Protocol 3, practice 5(b)).

A1.22 The Code also requires producers to “adopt common standards, concepts ... definitions ... and classifications” (Principle 4 practice 6).

A1.23 But this may not be sufficient, in reality, to enable statisticians to overcome barriers to increasing the harmonisation of classifications and definitions across administrative data holdings. The owners of administrative databases will, inevitably, require classifications that match the purpose of the particular type of administration; the Statistics Authority accepts this 'primacy' of purpose but also recognises that in some cases it would be possible to meet statisticians' needs without damaging the primary purpose of the system. It would be helpful if: (i) where administrative systems are used to produce statistics, statisticians were (as a matter of course) centrally involved in the design and operation of those systems; and (ii) there were to be a formal review of the implications for official statistics of the current use of different standards, definitions and classifications in administrative systems across government. Such an exercise might usefully be piloted initially.

Annex 2

Background information about the use of administrative sources for statistical purposes in the UK and internationally

- A2.1 The importance of administrative data for the production of official statistics is recognised in the European Commission Communication “Towards Robust Quality Management for European Statistics²⁵”. Intended to address the systematic problems revealed by Greece’s debt and deficit statistics, the Communication proposes that a more preventative rather than corrective approach be taken regarding the quality of government finance statistics. Specific measures that are proposed include further standardisation of public accounts and more regular assessments of the quality of data supplied by public bodies to National Statistics Institutes for the compilation of government finances²⁶. The Communication also proposes that Regulation 223/2009 - European Statistical Law - be amended to grant enhanced access to administrative records to professionally independent statisticians to enable them to compile statistics on national debt and deficit.
- A2.2 This drive to make better use of administrative data is closely aligned with a number of recent UK government policies and initiatives:
- In 2008 Richard Thomas and Dr Mark Walport conducted a review²⁷ of data sharing legislation with a remit to assess whether the operations of the *Data Protection Act 1998*²⁸ (*DPA 1998*) are appropriate, and to recommend an appropriate direction for the development of new data sharing legislation. The review concluded that there was a lot of confusion surrounding *DPA 1998* and that greater use could be made of the ability to share personal data safely, particularly for research and statistical analysis²⁹. The report made a series of recommendations, centred largely on clarifying and simplifying legal barriers, whilst maintaining privacy and security. Some recommendations specifically focussed on the use of data for statistical purposes and proposed measures to make it easier for academics and researchers to access government data in a secure environment. However, no specific proposals or recommendations were made to facilitate data sharing between Crown bodies for the production of official statistics.
 - The Localism Act³⁰ is intended to empower local people by devolving decision making powers to local authorities, neighbourhoods and communities. This would seem likely to stimulate demand for statistical information at the geographic level at which (local) decisions are made. Producing reliable statistics for small geographic levels based on sample surveys is typically costly, often too costly to be a realistic proposition.
 - The current Government’s drive to improve accountability through transparency (the Transparency Agenda³¹) requires that increasing amounts of government data are made publicly available.
- A2.3 Administrative data are also widely used in UK academic research, and the academic sector has taken a number of steps to maximise its use of administrative sources. The Administrative Data Liaison Service (ADLS) was set up to act as an intermediary between

²⁵ http://epp.eurostat.ec.europa.eu/portal/page/portal/quality/documents/COM-2011-211_Communication_Quality_Management_EN.pdf

²⁶ <http://europa.eu/rapid/pressReleasesAction.do?reference=IP/11/482&format=HTML&aged=0&language=EN&guiLanguage=en>

²⁷ <http://www.justice.gov.uk/reviews/docs/data-sharing-review-report.pdf>

²⁸ <http://www.legislation.gov.uk/ukpga/1998/29/contents>

²⁹ <http://www.justice.gov.uk/reviews/datasharing-intro.htm>

³⁰ <http://services.parliament.uk/bills/2010-11/localism.html>

³¹ <http://www.idea.gov.uk/idk/core/page.do?pageId=21407431>

data holders and academic organisations³², helping researchers to identify and gain access to administrative data. This service has stimulated new research methods which make use of administrative data, for example the use of capture-recapture techniques to estimate hidden populations³³. However, the remit of the ADLS is restricted to academia; there is no equivalent organisation whose remit extends to the public sector with a specific focus on the use of administrative data for the production of official statistics.

³² <http://www.adls.ac.uk/>

³³ <http://smm.sagepub.com/content/18/4/341.short>

Annex 3

Navigating the legal obstacles

A3.1 Despite the existence of useful guidance on the legal framework for data sharing issued by the Ministry of Justice (MoJ)³⁴, many departments have difficulty in navigating the legal obstacles. Difficulties include:

- Establishing the legal position when anonymised data are shared but the donor organisation has the key to identifying who those records relate to. A High Court judgment has clarified that anonymised information shared with others is not Personal Data³⁵ even if the donor organisation retains the fully identified version of the data. This is a recent judgment and government departments have yet to take it into account.
- Clarifying the roles of ‘Data Controller’ and ‘Data Processor’ when data are shared to support a purpose outside the scope of the donor of the data. There is an important distinction to be made between engaging a data processor to assist a department in its use of data, and providing data to another party for it to use for its own statistical purposes. Departments are usually more familiar with the former than the latter.
- Identifying the extent of implied powers to share data. Many departments consider that their implied powers have been eroded over time. In effect, as there has been so much legislation relating to data sharing, the opinion of some departmental lawyers is that the absence of a legislative authority for data sharing suggests that no implied power exists. MoJ has issued advice contrary to this widely held belief (which is an unintended consequence of legislation to enable data sharing), but many departments still do not recognise any common law powers to enable data sharing, even though such powers might be sufficient in some cases – for example, a one off study on a sample of administrative records might be considered ‘proportionate’.
- Establishing the meaning of ‘fairness’ in the first data protection principle³⁶. The re-use of administrative data for statistical purposes is not likely to be described clearly to, and actively endorsed by, members of the public who provide data about themselves in order to access government services. However, the *Code of Practice*, the *Statistics and Registration Service Act*, and the visibility given to data sharing for statistics can combine to make the re-use of personal information for statistical purposes fair even when users of government services have not been told directly about these uses of their data. For each Information Sharing Order made under the Act the Information Commissioner has been consulted. The issue of ‘fairness’ was addressed directly; the Information Commissioner’s view was that the use of the data would be fair even though the data subjects would not be informed directly about the re-use of their data for ONS’s statistical purposes.
- Complying with the right to a private and family life in the Human Rights Act. The important consideration here is to demonstrate that the re-use of personal information for statistics is necessary, proportionate, and not arbitrary. Proper planning, consultation on statistical work-programmes and, in particular, Privacy Impact Assessments can address this barrier.

³⁴ <http://www.justice.gov.uk/guidance/freedom-and-rights/data-sharing.htm>

³⁵ http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2007/wp136_en.pdf

³⁶ Processing personal data fairly and lawfully (principle 1).

Annex 4

Statements of Administrative Sources

A4.1 Protocol 3 (Practice 5) of the *Code* requires each organisation that produces official statistics (that are designated as National Statistics) to:

Prepare, in consultation with the National Statistician, a Statement of Administrative Sources which identifies the following.

- a. The administrative systems currently used in the production of official statistics.
- b. Procedures to be followed within the organisation to ensure that full account is taken of the implications for official statistics when changes to administrative systems are contemplated.
- c. Information on other administrative sources that are not currently used in the production of official statistics but have potential to be so used.
- d. Arrangements for providing statistical staff, whether inside the producer body or elsewhere, with access to administrative data for statistical purposes.
- e. Arrangements for auditing the quality of administrative data used for statistical purposes.
- f. Arrangements for ensuring the security of statistical processes that draw on administrative data.

A4.2 The remainder of this annex explains the value of these elements of Statements of Administrative Sources.

Protocol 3.5a - Scope

Statements should identify the administrative systems that are used to produce official statistics and should include both internal and externally owned systems. The purpose of this is to provide a 'stock-take' of the administrative systems currently used for statistical purposes.

Protocol 3.5b – Change management

Statements should explain how statistical teams are involved or informed when changes to administrative systems are contemplated. It is important that statisticians are aware of proposed changes so that they are able to determine whether this will affect the statistics, and take necessary action. This information is necessary to ensure that statistics produced using data from an administrative source meet the methodological and quality requirements of the *Code*.

Protocol 3.5c – Sources with statistical potential

Statements should identify any sources of administrative data – held within and outside of the producer organisation - that may be suitable for use in the production of official sources. This should also include information about how the suitability of data is determined. Publishing this information encourages the effective use of administrative data held by producers and promotes a unified statistical system. It also helps other producers of official statistics to maximise opportunities for the use of administrative data, facilitating the flow of information between producers of statistics (in accordance with Protocol 3, Practice 3 – maximise opportunities for the use of administrative data ... to avoid duplication - of the *Code*). This information is important for meeting Principle 6, Practices 3 and 5 of the *Code* - the identification and evaluation of administrative data sources as suitable alternatives to conducting surveys, thus reducing costs and burdens.

Protocol 3.5d - Access

Statements should specify the arrangements for providing statistical staff with access to data for statistical purposes. Such arrangements are necessary in order to meet the requirements of Protocol 3, Practice 3 of the *Code*, and to successfully maximise the use of administrative data.

Protocol 3.5e – Auditing Quality

Statements should detail quality audit arrangements. Quality auditing is necessary to ensure that statisticians (and custodians of administrative systems) are assured that the administrative operations and processes that lead to the recording of administrative data are suitable for the particular statistical purpose – such as providing consistent and comparable data. Such quality audit arrangements are an important step in meeting the methodological and quality requirements set out in Principle 4, Practices 1 to 3, and Principle 7, Practice 5, of the *Code*.

Protocol 3.5f - Security

Statements should explain the processes that producers follow in order to protect the security of administrative data that they use for statistical purposes. This information is important in its own right, but also enables the producer to demonstrate compliance with Principle 5, Practices 1, 2, 4, 5 and 6 of the *Code*.

Annex 5

The implications of the Open Data policy

- A5.1 The Government consultation document³⁷ *Making Open Data Real: A Public Consultation*, published in August 2011, built on the commitments initially set out in a letter from the Prime Minister to Government Departments in May 2010 (and subsequently reiterated and extended). He said,
- “Greater transparency across Government is at the heart of our shared commitment to enable the public to hold politicians and public bodies to account; to reduce the deficit and deliver better value for money in public spending; and to realise significant economic benefits by enabling businesses and non-profit organisations to build innovative applications and websites using public data.”
- A5.2 The consultation document made clear that a lot of new administrative data has already been put in the public domain and that much more is expected to follow³⁸. It is intrinsic to the policy that such data will be issued in a fairly raw form as soon as possible. It is assumed that the user community will find ways to make sense of it and that the benefits outlined in the Prime Minister’s letter will flow. The Statistics Authority offered its full support for Open Data in a letter to the Minister for the Cabinet Office³⁹ in October 2011, though noting that “‘data dumping’ can be inimical to transparency and good government. It is the task of government statisticians ... to produce official statistics which are intelligible, readily accessible, with objective and impartial commentary”.
- A5.3 There is more to both the Government’s plans for Open Data and the issues for statistics that those plans raise than can be addressed here. However, it is clear that the landscape in terms of available administrative data will be changing rapidly. Access to those data for statistical purposes may become easier, and the demand for analysis of it may become greater.
- A5.4 If ‘open data’ are to be used to allow the public to hold providers of public services to account, then quality and consistency will be critical. Nothing will undermine the credibility of open data faster than if it is used to challenge a public service provider only for the data to be shown subsequently to have been invalid or inconsistent between providers. There are documented cases of the quality of data (such as figures on ambulance response times) being adversely affected by their anticipated use to hold service providers to account – the so-called perverse incentive problem⁴⁰. It is necessary to be alert to the risk that the intended use of ‘open data’ may well add to the challenge of ensuring quality as well as increasing its importance.

³⁷ <http://www.cabinetoffice.gov.uk/resource-library/making-open-data-real-public-consultation>

³⁸ See for example <http://www.cabinetoffice.gov.uk/resource-library/open-data-measures-autumn-statement-2011>

³⁹ <http://www.statisticsauthority.gov.uk/reports---correspondence/correspondence/letter-from-sir-michael-scholar-to-rt-hon-francis-maude---open-data---11102011.pdf>

⁴⁰ A more recent example relates to crime maps: anecdotal evidence suggests that some people are reluctant to report anti-social behaviour to the police in case it impacts adversely upon house prices in their area.