

**ONS Virtual Microdata Laboratory:
Strategic Recommendations 2011-2015**

Tanvi Desai

(ESRC Placement Fellow)

London School of Economics and Political Science

March 2011

Contents

Executive Summary	P1
Recommendations	P2
1. Introduction	P4
2. Background	P4
2.1. Virtual Microdata Laboratory	P4
2.2. Secure Data Service	P5
3. Demand for VML Services	P6
3.1. Internal ONS demand	P6
3.2. External government demand	P8
3.3. Non-governmental demand	P8
3.4. Statistical leadership	P10
4. Strategic Recommendations	P11
4.1. Promote VML within ONS	P11
4.2. Centralise management of VML	P11
4.3. Position VML as a key corporate asset	P12
4.4. Fund VML as a key corporate service	P12
4.5. Use VML to distribute access not data	P13
4.6. Preserve staff expertise	P13
4.7. Increase engagement with OGDs	P14
4.8. Develop VML as the basis for a GSS-wide service	P14
4.9. Governance	P15
5. Conclusion	P15
References	P18
Annex A: Statistical Leadership – Organisations consulting MAUS	P20
Annex B: Glossary of Acronyms	P20
Acknowledgements	P20

Executive Summary

- Every year tens of millions of pounds are spent by government and funding bodies on collecting data. The majority of this data is only ever published as headline (macro) statistics, with the underlying microdata underexploited or completely unused.
- There is a strategic focus UK wide on maximising returns to existing data, investigating the potential offered by administrative data, and minimising the cost and of data collection and burden on respondents. All this must be done while ensuring that data confidentiality is preserved.

'Effective use of information is absolutely central to the challenges facing the Government today – whether in improving health, tackling child poverty, or protecting the public from crime and terrorism. Those in public service need to keep that information secure, in order to build public confidence. This is essential to underpin greater data sharing to deliver personalised services and make us more effective.' Sir Gus O'Donnell (CO 2008)

- The Data Sharing Review (Thomas and Walport 2008) recommends that to preserve data confidentiality *"safe havens" should be developed as an environment for population-based research and statistical analysis'*.
- The ONS Virtual Microdata Laboratory (VML) is a 'safe haven', which the OECD called *'one of the most innovative research efforts in the public sector across the 30 OECD member countries'* (VML 2007). Established since 2004 the VML's primary role has been as a service supporting policy based research using confidential microdata. The majority of VML users are academic researchers, many working on government funded policy analysis.
- Changes in the UK data infrastructure mean that the majority of academic researchers will transfer to the ESRC funded Secure Data Service from April 2011. With the financial pressures on all government departments there is a need to assess whether sufficient demand remains to justify ONS' continued investment in VML, and if so how that investment should be focused to ensure maximum possible value for money.
- There is a strong demand from within ONS for an infrastructure that allows sensitive data to be manipulated in a secure environment. This is particularly important for several core government initiatives being led by the ONS where highly sensitive data from a range of sources are brought together to be matched at the respondent level e.g. the Migration Statistics Improvement Programme, Beyond 2011. It should also be noted that, as SDS will initially only be able to provide a service for academics, if ONS staff wish to continue research using sensitive microdata then the VML or an equivalent will be necessary to maintain data security.
- In the VML the ONS has an established, internationally respected, relatively low cost¹ service that supports many of the ONS' strategic and business aims. In addition to providing an infrastructure in which existing data can be fully exploited for policy based research, the VML plays a key role in improving official statistics, whether as an infrastructure for major ONS projects to store and combine data sources or as an interface between government and

¹ In comparison to similar services internationally

academia enabling ONS to harness top quality academic expertise. The service also provides statistical leadership in data security, advising institutions nationally and internationally.

- It is highly unlikely that discontinuing support for the VML would lead to cost savings for the ONS, as even in the short term core ONS projects such as the MSIP will require a replacement to be provided in order to maintain data security. Though it is clear that the VML is not currently positioned to offer the best possible return for the ONS' investment, the service is only in need of a small amount of refocusing to become a key ONS asset. This would be far more cost effective than attempting to develop an equivalent service from scratch.

Recommendations

1. Promote VML within ONS

For the past six years the VML has largely focused on providing a service to the academic community. With the foundation of the ESRC Secure Data Service (SDS) the majority of academic researchers will no longer use the VML. This is an opportunity for VML to refocus on providing a service for the ONS at a time when all units must prove their value for money to the institution funding them.

VML should increase promotion within the ONS, through advertising, workshops and individual consultation with senior staff to see how the service could be tailored to meet their needs.

2. Centralise management of VML

The current policy whereby internal ONS projects all manage their own 'instances' of the VML leads to a variable approach to data security. In order to establish clear common standards and procedures, access and security management should be handled by a central management team.

3. Position VML as a key corporate asset

As it is currently positioned within the Economic Interpretation Division of the Economic, Labour and Social Analysis Directorate the VML is largely viewed as a research support service by those who are aware of it. However this significantly underestimates its potential value to the ONS. Positioning the VML as a corporate service would raise the profile of the service, increase VML staff authority to implement data security standards, and decrease the possibility of duplicate services being developed within ONS.

4. Fund VML as a key corporate asset

With the departure of the majority of academic researchers to the SDS, the main role for the VML is likely to be in supporting the development of official statistics by providing a secure environment for processing sensitive data. While there will still be demand for the service for data analysis for research purposes it no longer makes sense for the Economic Interpretation Division to bear the full burden of funding the service. Therefore it is recommended that the VML be funded centrally as a corporate service.

5. Use VML to distribute access not data

A recent agreement to allow VML terminals to be rolled out across the GSi will allow VML to distribute access to government users of ONS data. This will remove the need for ONS to post large amounts of highly confidential data to government officials on CD, and support ONS' (and UK government's) aim to disseminate access instead of data.

6. Preserve staff expertise

One key area of risk for VML is the loss of staff expertise. Data security in particular is vulnerable to staff turnover. Staff who have the appropriate skills to undertake manual statistical disclosure control to the level required by VML are rare. Where possible ONS should allow experienced staff to stay in post beyond the standard 2 year period.

ONS should ensure that there is an effective knowledge transfer mechanism in place, so that the expertise is not lost when staff move posts. Without the expertise developed by the MAUS team ONS will have to cede leadership in this area to the staff at SDS leaving little expertise within the GSS.

7. Increase engagement with other government departments

The MAUS team has already been consulted by a number of UK government departments on data security issues. There is potential for ONS to take leadership in this area by making MAUS expertise explicitly available to other government departments, either via workshops, GSS policies or documents, and/or individual consultations.

Additionally the MAUS infrastructure could be offered as an archiving and data analysis facility, on a cost sharing basis, to other government departments.

8. Develop a GSS wide 'safe haven'

The way to maximise the returns to investment in the VML is as the basis for a GSS wide archiving, development and analysis facility. Not only would such a facility spread the cost of the service across government as a whole, negating the need for multiple department specific services, but it also offers the potential for significant savings in the cost of data collection. Bringing government data sources together in one place would make it easier to maximise returns to existing data and to audit what is already available to minimise duplication. Such a strategy is also likely to lead to a reduction in the respondent burden.

9. Establish a steering committee

A service with the potential sensitivity of the VML should have an effective advisory or steering committee. A steering committee can advise on strategic planning, knowledge transfer, adoption of standards, quality control and can act as the final arbiter in disputes over conduct and access. While the agreement between SDS and VML for mutual peer review of output will assist in maintaining standards, a senior body overseeing operations will help to establish the VML as an authoritative entity as well as enabling ONS to oversee operations more effectively.

1. Introduction

The Office for National Statistics' (ONS) Virtual Microdata Laboratory (VML) has formed a significant part of the UK data infrastructure for the past 7 years, providing secure access to confidential microdata resources for policy relevant research and enabling the ONS to ensure that data resources collected at significant cost are exploited as fully as possible.

With the major changes in the UK data infrastructure and approaching financial pressures on all government departments, the challenge is to assess whether there is sufficient demand for the VML to justify ONS investment, and if there is, to ensure that its operations are directed in such a way that the maximum value for money is achieved.

To assist in addressing this issue, the ONS and the Economic and Social Research Council (ESRC) joint funded a Placement Fellowship² based at the VML to bring in external expertise to contribute to a strategic and operational review. The post comprised three major tasks, in addition to the development of the VML strategy, the Fellowship tasks were to manage and review VML operations, and to facilitate the development of the new ESRC funded Secure Data Service, more details of which can be found below.

This paper will briefly outline the VML and the SDS; it will then detail areas of demand for VML services which the author has observed during the period of the placement; finally the paper will make suggestions as to how the VML can focus its future strategy to ensure that it provides the most secure, efficient and cost-effective service going forward to 2015.

2. Background

2.1 Virtual Microdata Laboratory

The VML is the ONS' 'safe haven' and was established in 2004 to provide a secure solution for research access to confidential business microdata. The VML is managed and supported by staff in the Microdata Analysis and User Support (MAUS) branch based in the Economic, Labour and Social Analysis Directorate (ELS).

The VML provides on-site access to confidential data³ at ONS offices in London, Newport, Titchfield, Glasgow and Belfast. At these sites 'Approved Researchers'⁴ can access highly detailed microdata for research purposes. While at a VML site researchers have full access to data they have been 'approved' to use, but they cannot remove any data from the ONS premises⁵, any output they wish to remove must be submitted to the MAUS team to be manually checked for disclosive data. Only when output has been confirmed not to contain any disclosive information is it forwarded to researchers for inclusion in publications.

Since its foundation in 2004, the VML has proved to be a very successful service showing a steady increase in research use by academia, the ONS and other government departments. In the past two years there has been additional demand in VML resources from core ONS projects that require a secure setting for their data resources, for example the Migration Statistics Improvement Programme.

² <http://www.esrc.ac.uk/funding-and-guidance/collaboration/knowledge-exchange/opportunities/index.aspx>

³ Since 2005 social data has also been made available via VML

⁴ As defined by the Statistics and Registration Services Act

⁵ For details of operational and technical controls see Ritchie 2011a

The MAUS team's expertise has led to them playing a leading role in setting standards in secure microdata access both nationally and internationally. The HMRC and DWP have both consulted the MAUS team when considering the development of secure data access solutions, with the HMRC adopting VML procedures. The new ESRC funded Secure Data Service is also based on the VML model.

VML has been a strong presence for the ONS, not only nationally, but at Eurostat and in other EU funded projects, enabling the UK to have a significant influence on the future of the data access infrastructure in Europe. In fact the VML influence has been so successful that a recent ESSNet⁶ project to establish a standard for secure access to data in Europe has largely adopted the VML statistical disclosure control (SDC) and researcher training principles (Brandt et al. 2010) (Brandt & Eilsberger 2009)⁷.

The VML costs approximately £350,000 per year to run⁸. This contrasts very favourably with other similar services throughout Europe and the US⁹. This has largely been due to the innovative, principles based approach to data security that requires a dialogue with data users (for more details see (Desai & Ritchie 2009)). This approach not only enables the ONS to achieve a high level of satisfaction among users, but also makes it possible for ONS to leverage researcher expertise in the development of SDC procedures. The principles on which VML is run are strongly aligned with the ONS strategic aim *'to engage constructively and more broadly with our customers'* as MAUS staff are in daily conversation with users about their research and data needs.

The key challenge faced by VML at this time is that the focus of the service has been primarily on providing access to academic researchers¹⁰. With the establishment of the ESRC's Secure Data Service the demand among academic users for VML resources will fall considerably, therefore one of the main issues to address in any future VML strategy is how to preserve the expertise accumulated by the VML while redirecting the focus of VML operations to where it is needed.

2.2 Secure Data Service

The Secure Data Service (SDS)¹¹ was funded by the ESRC following a recommendation in the first National Data Strategy (Elias 2006). The SDS will provide secure remote access to confidential microdata deemed too sensitive to release in any other form. The Service is hosted by the UK Data Archive¹² (UKDA) at Essex University, which has provided access to microdata generated by government and researchers for over 40 years.

In November 2008 the then National Statistician, Karen Dunnell, wrote to Ian Diamond the Chief Executive of the ESRC supporting the establishment of the SDS, and committing ONS to expanding its

⁶ <http://epp.eurostat.ec.europa.eu/portal/page/portal/essnet/introduction>

⁷ A list of organisations that have consulted MAUS can be found in Annex A

⁸ Not including ESRC contributions to academic access costs, the VML raises on average £25,000-£35,000 per year (£15-£25,000 in access charges, and £10,000 for analytical work).

⁹ In 2008 the Swedish equivalent of VML cost €1m per year to run, the Danish cost £600,000 per year in 2003, the US census bureau datalab employs 40 staff compared to 7.7 FTEs in the MAUS team, while the budget for the Netherlands system has been said to be an 'order of magnitude' higher than ONS.

¹⁰ For which ESRC has provided grants to ONS totalling £230,000 over the past 4 years

¹¹ <http://securedata.ukda.ac.uk/>

¹² <http://www.data-archive.ac.uk/Introduction.asp>

partnership with the UKDA. The aim was to make confidential microdata available to researchers through the SDS, with a particular focus on the collection already available for research access in the VML.

In order to ensure that SDS operates to ONS standards, SDS training and security procedures are based on those of the VML, and there is an agreement in place between ONS and UKDA to carry out mutual peer review of the output released by each service in order to maintain high standards of security. In addition to operating to VML standards, the Secure Data Service has at ONS' request secured ISO27001¹³ certification and hired an external company to undertake internal and external penetration testing.

During the first phase, access to data via SDS will be restricted to academic users, and researchers working on ESRC funded projects. The long term plan is to expand access to SDS beyond academia, and suitable mechanisms are being investigated in advance of the next SDS funding round in 2012. However, considering inevitable funding constraints over the next few years, contributions from bodies other than the ESRC are likely to be necessary to make this possible.

As the ONS will implement a daily charge for academics accessing VML after the 1st April 2011, expected to be in the region of £85, considerations of cost and convenience will mean that the majority of VML's academic users (over 80% of current users) will transfer to the SDS. While this will reduce the need for ESRC to fund academic access to VML, and will address a situation where ONS has to some extent been subsidising academic research, it also means that the VML will lose almost all of its core users, necessitating a change of focus from primarily supporting academic research to supporting internal ONS projects, and providing statistical leadership in secure data access.

3. Demand for VML Services

There is a clear demand for the VML: for the MAUS team's expertise as well as the VML infrastructure itself, both from internal ONS staff, and from the wider data community.

The VML supports many of the ONS' strategic aims including the aim '*to exploit data sources for analytical potential*' (Whitstone 2010). In this section existing areas of demand will be explored, these will then contribute to forming the overall strategic recommendations.

3.1 Internal ONS demand

The VML plays a key role supporting a number of ONS strategic and business aims and objectives in its' position as the ONS 'safe haven'.

The demand for a VML style service in which data can be analysed and manipulated securely is clear from a 2010 business case submitted by the Administrative Sources unit for £1.2m over five years to develop a secure infrastructure very similar to the VML.

¹³ ISO/IEC 27001:2005 - Information technology -- Security techniques -- Information security management systems – Requirements is an Information Security Management System standard published by the International Organization for Standardization <http://www.iso.org/iso/home.htm>

The Administrative Sources unit is leading the Migration Statistics Improvement Programme¹⁴ (MSIP) which involves combining administrative and other data from a range of sources including other government departments (OGDs) to improve the measurement of migration internally and internationally. As these data are to be matched, they are necessarily highly disclosive and it is crucial that both the respondents and the government departments supplying the data are protected. The MSIP programme made use of the VML during their pilot phase and contributed to an upgrade of the VML IT systems to ensure storage and processing was sufficient for the project's needs. Since the upgrade the Administrative Sources unit have abandoned plans to develop an equivalent service and started to transfer significant amounts of data to VML, further highlighting both the value of VML, but also the need to establish robust policies for the use and governance of the VML.

Other departments within ONS making use of VML for data processing include Methodology, Prices, and the Census.

Requirements for a secure area for data matching underpin major ONS projects aimed at improving the quality of official statistics such as the MSIP, Beyond 2011, and potentially the Well-being Index; however there is no justification for a number of parallel systems providing almost identical services. Using the established VML infrastructure as a starting point in developing a single corporate asset would allow the ONS to

- continue the work that has been started by the MAUS team in developing and promoting standards for SDC and for 'safe havens'
- allow other units to draw on the expertise developed by the MAUS team
- provide all divisions of ONS with an established, secure, centralised infrastructure in which to process sensitive data
- increase data security by creating one central, auditable point for data matching
- archive microdata for future exploitation that might otherwise have been lost
- ensure the best possible use of resources at a time when financial savings are vital.

It should also be noted that as SDS will not be able to provide data access to government researchers¹⁵, a VML-type infrastructure will be necessary for ONS staff to be able to continue to analyse confidential microdata for research purposes. There are currently 24¹⁶ ONS research projects that make use of the VML infrastructure and this is likely to increase if the MAUS team refocuses to engage more with government.

There is demand for the VML within ONS, and if ONS were to discontinue support of VML it would still be necessary to develop a similar service, in particular to support major projects that require a secure environment for data matching. Therefore the ONS has the choice to build on the existing infrastructure and utilise the expertise developed within the MAUS team, or to start from scratch in developing a new service which would require a far greater investment.

¹⁴ <http://www.ons.gov.uk/about-statistics/methodology-and-quality/imps>

¹⁵ SDS will only be available to academic researchers see p6

¹⁶ Not including projects accessing the longitudinal study

3.2 External Government Demand

Requirements for government departments to prove that they are generating the best possible return for their investment in data collection mean that OGDs are increasingly looking into ways to make their data available for research purposes. The Data Sharing Review (Thomas & Walport 2008) recommended that *“safe havens” should be developed as an environment for population based research and statistical analysis’*. As the VML is an established ‘safe haven’ a number of government departments have already consulted with the MAUS team, and are investigating the possibility of setting up their own equivalents to VML.

The MAUS team’s expertise can be capitalised on in *‘setting clear common standards and procedures’* as recommended by the ‘Report on Data Handling Procedures in Government’ (CO2008).

The strategy that offers the best possible return to any investment in VML is for the service to form the basis of an ONS led, government-wide ‘safe haven’, acting as a central repository, archive and data analysis environment for the government as a whole. This ‘joined up’ approach would allow ONS to take statistical leadership of a key GSS resource; it would allow government departments to pool resources in the development of secure data access; and it would enable the GSS to establish a robust standard for UK ‘safe havens’. If necessary, the VML architecture would allow data and (less desirably SDC) to be distributed, so that OGDs could store data within their departments (and potentially check output), while the system would be managed by the ONS. A likely additional benefit of bringing government data resources together would be a more coherent overview of the data available within the UK, enabling assessment of where the duplications lie, thus reducing the burden on respondents as well as the cost of data collection.

The ONS Statistical Policy Committee has agreed to allow VML terminals to be rolled out across the GSi, with the Welsh Assembly Government, and the Department for Business Innovation and Skills (BIS) already setting up terminals. This will allow VML to distribute access to ONS data, removing the need to post large amounts of highly confidential data to government officials on CD, and supporting ONS’ (and UK government’s) aim to disseminate access instead of data. Using VML as a solution to disseminate access would

- significantly improve ONS’ data security
- remove the need for untraceable CDs containing confidential data to be in circulation
- enable ONS to increase their knowledge of what their data is used for
- reduce the amount of time data custodians spend in generating data to distribute.

3.3 Non-governmental demand

Over the past 6 years the VML has offered an unprecedented opportunity for government and academic researchers to collaborate on policy oriented research using highly confidential data, in particular firm level data. The VML has also enabled the ONS to leverage academic expertise in improving statistical outputs. Policy areas which have benefited from academic research in the VML include: the measurement of intangible investment (Marrano et al 2007); the climate change levy

(Martin et al 2009); ICT impact on the UK economy (Bloom et al, 2007) among others¹⁷. In fact Professor John Van Reenen¹⁸ (2010) stated that

'The VML is extremely cost-effective since researchers actually pay ONS to use the VML through the ESRC. If a government department wanted to obtain the sort of high quality information generated by VML research they would have to pay very high costs in-house or from consultancies. Since academics are motivated to produce this research for journals which determines their career success, their time is effectively costless for the government. Since the people working on the data are generally PhD-trained highly able individuals, this is an amazingly cost-effective way of generating analysis on leading UK databases.'

The new research areas opened up by the availability of business and small area data are unlikely to cease if the VML is no longer available, particularly since some of the research strands are key to government policy. This means that researchers will continue to request access to confidential data.

Following the Statistics and Registration Services Act 2007 (SRSA) there is a *'formal and permanent recognition by the State that one of the roles of an NSI is to provide research access to data in a way that safeguards confidentiality'* and that *'any source of data, in any detail, is in scope'* (Jackson 2008). The VML allows the ONS to fulfil this role by providing an infrastructure that is tried and tested for the release of confidential microdata for research. This means that the ONS can release data to researchers via the VML with the confidence that the infrastructure is fit for purpose and the administration and statistical disclosure control of the data is being managed by experienced staff.

It has already been noted above that a VML type infrastructure is vital to ONS and government activities to improve official statistics via administrative data linking (such as MSIP, Beyond 2011 etc). As the ONS will not initially allow data in SDS to be linked at the level of the individual respondent it is likely that there will be a similar demand for VML from non-governmental researchers who wish to undertake data linking.

The VML can also be used as a test bed for newly released data. For example if an academic researcher requests access to previously unreleased microdata, the ONS might wish to make that release via an internal system over which they have full control in order to monitor use of the data. In time as it was proved that use of the data was producing valuable results while maintaining respondent confidentiality the data might then be passed to SDS. Such a strategy allows the ONS to fulfil its obligations under the SRSA to support innovative research and to ensure maximum possible exploitation of the data, while removing the need for data to move outside the Government Secure Intranet.

VML also has a key role to play in allowing UK statisticians and researchers to access European microdata. Negotiations among European member states are leading to the possibility of sharing microdata across borders. If these continue in a similar direction then the ability of any member

¹⁷ For more information on research undertaken at VML see the MAUS quarterly bulletins <http://www.ons.gov.uk/about/who-we-are/our-services/vml/about-the-vml/outputs-and-publications/index.html>

¹⁸ Director of the Centre for Economic Performance at the London School of Economics http://cep.lse.ac.uk/_new/staff/person.asp?id=1358

state to receive microdata from other countries will be based on the level of infrastructure in place at that member state's national statistical office. Therefore the UK are currently very well placed to receive European microdata as the VML is taken as an example of best practice across Europe. However without the VML the UK will not have the infrastructure to receive data from many member states. As European law currently only allows data to be transferred to a national statistical office, it will not be possible to substitute the infrastructure provided by the SDS for the VML in this case. This would leave UK government and researchers in a weakened position when carrying out Europe wide comparative research.

There is a strong demand for potentially sensitive microdata among academic and non-government researchers who wish to undertake innovative research. The VML enables the ONS to provide access to such data while maintaining the data security. The infrastructure also acts as a point of contact between researchers and the ONS giving ONS the opportunity to profit from research undertaken within the VML; from researcher expertise in improving official statistics; and to '*engage constructively and more broadly with our customers*' whether those customers are the researchers themselves or OGDs commissioning policy based research that requires confidential data.

3.4 Statistical Leadership

The expertise accrued in the VML has enabled the staff to provide statistical leadership in a number of areas including

- the ONS strategy on 'Provision of ONS data for analysis' (Ritchie 2011b)
- the promotion of an understanding of data security among data users
- the development of a standard for statistical disclosure control and researcher training
- the development of European strategy for access to confidential data
- advice for international organisations concerned with data security and data access.

VML has also provided expert comment on the ONS' Research Data Access strategy developed by Legal Services.

Felix Ritchie's principles of secure data access: *safe projects, safe people, safe data, safe setting* first explored in 2002 (Ritchie 2002) and published formally in 2004 (Ritchie 2004) have since become a standard for the field, and are commonly cited in papers and presentations internationally.

The VML training course for users has fostered a greater understanding of data security issues among researchers. Previous data release strategies that relied on 'safe data' and an 'us and them approach' (Desai & Ritchie 2009) led to a culture where researchers were not required to have much understanding of data security issues beyond abiding by a few basic rules (not identifying individuals, not sharing licensed data etc). With the growing strategic focus on maximising the return to investment in data collection, and in maximising the use of administrative data, ONS and other data collectors are increasingly looking at the potential offered by more detailed, hence sensitive data. This means that relying on the 'safe data' approach is no longer practical and there is a need for leadership in introducing data security standards to researchers and staff in government and academia. VML has been providing this leadership and, alongside European and UK organisations such as ESSNet and the UK Data Archive, has been rolling out both the VML researcher training

programme and the VML statistical disclosure control methodology as the foundation of new standards in data security.

4. Strategic recommendations

Recent increased demand for VML services from within ONS, in particular in relation to key strategic projects such as the Census, Beyond 2011, and MSIP demonstrates that VML has a role to play within ONS. Discontinuing support for the VML is likely to be financially costly as these core projects will still require an equivalent infrastructure to maintain data security, obliging ONS to develop a new service from scratch.

The advent of ESRC's Secure Data Service can be seen as an advantage to the future strategy of the VML. The transfer of academic researchers to the SDS removes the core burden on VML staff time and allows VML to strategically refocus towards serving the needs of the ONS and other government departments.

This section will make recommendations on where VML expertise and infrastructure should be focused to ensure maximum possible return to the ONS' investment.

4.1 Promote VML within ONS

For the past six years VML has largely focused on providing services for academic researchers. There seems to have been little official promotion of the service within the ONS. However through word-of-mouth alone, ONS staff have come to account for 13%¹⁹ of active projects within the VML. In addition to these one-off projects, the VML is also a key infrastructure for the Administrative Data project, as evidenced by the Population and Demography section (of which the Administrative Sources unit is a branch) committing £150,000 in the financial year 2009-10 to upgrading the VML to enable storage of the data required for projects involving data matching.

In addition to the Administrative data project the VML also supports, the Census, Methodology, and Prices, and will be needed by ONS projects that require sensitive data from a range of sources to be brought together for linking such as Beyond 2011.

It is clear that a 'word of mouth' method of promoting the VML has led to the infrastructure being underexploited within the ONS, and in the worst case scenario could lead to services being developed that replicate those offered by VML due to a lack of awareness of the service across ONS. A major priority for the MAUS team should be to promote the VML across the ONS, and to liaise formally with senior staff to see how VML can support their programmes of work.

4.2 Centralise management of VML

To support core ONS projects with a need for significant resources the VML is partitioned into what are referred to as 'instances': dedicated secure spaces within the VML structure. The current practice is for decentralised management of these instances, allowing each project to control access, and security within their area. This has the disadvantage that security practices across 'instances' vary widely and it is fairly clear that the standards outside MAUS controlled 'instances' often fall

¹⁹ This does not include ONS core programmes or delegated instances for ONS methodology work, administrative data etc.

below established VML levels. Examples of bad practice, such as not removing the accounts of staff who no longer work on the project (or even at the ONS), complaints that maintaining a record of who has access to data is 'unreasonably bureaucratic' and requirements for a large number of administrative accounts are clearly not within the ONS' code of practice, let alone up to the standard of the VML's central administration. When handling highly sensitive data from a range of sources, including other government departments, it is crucial that ONS be seen as not only utilising, but establishing standards of best practice, and that these standards are consistent across the organisation. Therefore a key recommendation would be that the MAUS staff manage all instances of the VML to ensure that ONS can support access to sensitive data to a consistent and high standard.

4.3 Position VML as a key corporate asset

One major obstacle to MAUS staff being able to implement standards across all instances is the perceived lack of seniority of the service within ONS. The VML is currently positioned in the Microdata Analysis and User Services branch of the Economic Interpretation Division (EID) of the Economic, Labour and Social Analysis directorate and as such is not viewed as a core ONS service.

In the VML the ONS has an established, proven and internationally respected infrastructure for the management and analysis of confidential data. To enable the ONS to fully exploit the VML's potential as a corporate asset the service should be moved from within the Economic, Labour and Social Analysis directorate (ELS) to be positioned more centrally. As long as the service is buried within ELS there will be a tendency among the rest of the ONS (those that are in the position to discover its existence) to view VML as a research support service, which significantly underestimates the potential of the infrastructure that has been developed.

A more central position within ONS could enable VML to

- increase visibility within the organisation
- develop a role as a core ONS service providing infrastructure for major strategic projects
- establish the authority necessary to impose standardised high quality security controls thus improving data security within ONS
- increase the likelihood of cost savings by reducing the risk of requests for funding to create similar services
- support the ONS strategic aim to disseminate access not data

4.4 Fund VML as a corporate service

A more centralised position within the ONS would also enable the cost of the VML to be shared more equitably across the ONS. The VML is currently funded entirely from the budget of the Economic Interpretation Division despite the fact that much of its core value to the ONS is as an infrastructure for the development and improvement of official statistics. Positioning the service as a corporate service could enable costs to be shared across departments that make use of the data, or preferably to be fully funded centrally as a service for the ONS as a whole.

4.5 Use VML to distribute access not data

There is a strategic aim across UK government to tighten the security of sensitive data by distributing access to data rather than data itself. In the VML, the ONS already has an established infrastructure that allows them to meet this aim.

The ONS currently posts sensitive data on CD to government officials including regional statisticians. Once data has been posted out ONS loses control over what is subsequently done with that data. No matter how strict access conditions are, data on CD will never be as safe as data stored and accessed in a safe haven. Following permission from the Statistical Policy Committee for VML to be rolled out across the government secure intranet, ONS has the option to use the VML to replace the distribution of data on CD for government officials.

4.6 Preserve staff expertise

One key area of risk for VML is the loss of staff expertise. Data security in particular is vulnerable to staff turnover. The VML's use of a 'principles' based approach to data security instead of a traditional 'rules' based approach has many advantages as detailed in (Desai & Ritchie 2009) not least of which is a significant cost saving over the traditional method. However, it requires experienced staff with a good understanding of econometric and statistical analyses, of research practice, and most importantly of how the VML SDC principles are implemented. There is a danger that a high staff turnover will leave the VML without the expertise to implement the SDC rules to a level that will ensure data security. The current situation where the MAUS team looks likely to lose all staff, apart from the administrators, within a short space of time is likely to leave the security of the VML seriously flawed unless action can be taken to retain the grade 6 and/or to secure knowledge transfer to a level where the ONS and MAUS staff can be confident that they can maintain standards.

There is an urgent need to implement a programme to preserve and transfer the expertise developed by the MAUS team. Ideally all MAUS staff should also have access to a senior member of staff with a thorough understanding of the principles based approach to provide authoritative advice in case of uncertainties or disagreements. If ONS cannot preserve this knowledge then the only alternative is to allow the Secure Data Service to take full leadership in this area, leaving ONS data security vulnerable, and VML no longer fit for purpose as a service for research data analysis as it will no longer be possible to carry out statistical disclosure control of outputs effectively.

Preserving the VML's 'institutional memory' in the shape of administrative expertise is also very important. MAUS administrators maintain contacts with data custodians, IT specialists, and users, understand the structure and procedures necessary for the efficient functioning of the service, and provide an underpinning for the service, ensuring performance targets are met. Administrative consistency is particularly important at a time of high staff turnover.

One issue that should be addressed when appointing staff to the MAUS team is the balance between research expertise and administrative expertise. The current policy of having support staff with a research background has been very successful. The policy has facilitated communication between MAUS staff and researchers, both because having a research background enables MAUS staff to grasp a user's problem more effectively, but also because researchers tend to view them as 'one of

us'. There is however a disadvantage to having MAUS staff who are too focused on research, that is a tendency to view the administrative procedures as uninteresting or in some way 'beneath them' leading to lax output checking and SDC measures. There is also a danger with staff who are too connected to former (or potential future) academic colleagues that there may be a conflict of interest that affects their neutrality when it comes to disclosure control. This problem is likely to be significantly reduced by the removal of academic staff to the SDS, however when appointing new staff, the candidate's research experience should be balanced by a genuine interest in and commitment to data security.

4.7 Increase engagement with OGDs

The MAUS team has already taken a role in advising other government departments on managing their secure data access. For example DWP have consulted with MAUS on a number of occasions, and HMRC are in the process of developing a 'safe haven' based on the VML procedures and principles, while BIS has an 'instance' already established within VML. The VML procedures are rapidly becoming accepted as a national as well as Europe-wide standard for researcher management and statistical disclosure control. The number of international speaking invitations received by the MAUS grade 6 over the past two years²⁰ is clear proof that the expertise developed by MAUS is highly prized internationally.

There is scope for ONS to further exploit and promote the statistical leadership role MAUS can play, providing guidance and rolling out internationally respected standards for the Government Statistical Service as a whole. It is recommended that workshops and policy documents are developed, and consultations offered to promote and harness MAUS expertise for the GSS.

4.8 Develop VML as the basis for a GSS-wide service

At a time when economies of scale are vital, there is also the potential for ONS to take the lead in developing a government wide 'safe haven'. The current plan for individual government departments to develop their own solutions to confidential data access is very costly, and necessitates duplication of efforts, roles, and resources across government. As well as the financial implications there are staffing implications to this approach, as the challenge of finding an adequate number of suitably qualified staff to carry out SDC for all government departments to the level required is likely to be significant. One way to address this issue would be for the ONS to take a leadership role in expanding the VML to support the GSS more widely, bringing together data sources from different government departments. This would have a number of advantages

- Easier to develop and maintain UK wide standards
- A financial saving for government departments
- Increased data security across government
- In line with government strategy to distribute access rather than data
- Less need to transfer data between government departments (increased security)
- An archive of government microdata sources that might otherwise be lost
- Easier audits of what government data is used for

²⁰ For a list of organisation that have consulted MAUS see annex A

- Collecting a range of data sources in one place will facilitate government strategy to maximise returns to data collection, and to minimise the respondent burden by allowing data to be audited and combined more effectively

4.9 Governance.

There is currently no effective governance for VML beyond the standard ONS hierarchy. While an advisory committee was set up when the VML was founded this has since lapsed, and all decisions and responsibilities for the service fall entirely on the Grade 6 MAUS manager. For a service that is as potentially sensitive as the VML this is not an acceptable position.

A steering committee should be formed to oversee the VML with a remit to contribute to strategic planning, knowledge transfer, standard setting, as well as acting as a final authority in disputes over conduct and access.

It is recommended that the steering committee membership includes a senior member of the ONS with responsibility for the VML, the MAUS manager, a senior member of the ONS legal team with a connection to the Microdata Release Panel, and the director of the Secure Data Service. The committee might also include senior representatives from OGDs who deposit or access data in VML, a senior member of the academic community, and a representative of a similar service from another national statistical office.

5. Conclusion

The UK has long been a leader in the use of microdata for evidence based policy. The 2008 International Benchmarking Review of UK Economics stated that

*'First and foremost, the research achievements of UK scholars are exceptional by world standards; the UK economics profession is more prominent than any other country's except for the United States. UK scholarship has been very influential in a number of important fields, such as labour economics, public economics, and economic development, and it has attained world leadership in microeconometrics.'*²¹ (Helpman 2008)

This would not have been possible without high quality data and innovative approaches to data access. However despite this, much of the data collected, at the cost of ten of millions of pounds to government and survey respondents²² every year, is used only to create macro level statistics. Though the marginal cost of making data available for research is very small compared with the cost of collecting data, confidentiality concerns, lack of suitable infrastructure or lack of incentive mean that most of the microdata collected is either underexploited or put to no further use at all. As the Statistics Authority Code of Practice states that *'cost burden should not be excessive and should be assessed relative to the benefits arising from the use of the statistics'* there is a clear incentive for ONS to ensure maximum possible use of, and therefore benefits from, their data.

²¹ http://www.esrc.ac.uk/images/IBR_economics_tcm8-4664.pdf

²² In 2009/10 the cost of social surveys wholly or partly funded by the ONS alone was over £30m with the compliance costs coming to over £10m (ONS 2010a). Longitudinal surveys cost government over £17m in the same period (ONS 2010b)

In the VML the ONS has an established 'safe haven' for making confidential data available for research as recommended in the Data Sharing Review (Thomas & Walport 2008). In addition to providing an environment in which existing data can be fully exploited for policy based research, the VML plays a key role in improving official statistics, whether as an infrastructure for major ONS projects to store and combine data sources or as an interface between government and academia, enabling ONS to harness top quality academic expertise. The VML also provides statistical leadership in data security, advising institutions nationally and internationally.

The VML not only underpins major ONS projects but helps ONS to fulfil Strategic Aims and Objectives including

- *To be recognised for our statistical leadership*
- *To engage constructively and more broadly with our customers*
- *To manage information risks effectively*
- *To comply with the Code of Practice for Official Statistics*
- *To plan for innovation*
- *To encourage a culture of innovation*
- *To exploit data sources for analytical potential* (Whitestone 2010)

It is highly unlikely that discontinuing support for the VML would lead to cost savings for the ONS, as even in the short term major ONS projects such as the MSIP will require a replacement to be provided in order to maintain data security. However it is clear that the VML is not currently positioned to offer the best possible return for the ONS' investment.

Firstly the service should be positioned more centrally within the ONS. As long as it is within EID it has low visibility within the organisation and there is a tendency among ONS staff to view it as a research support service, which leads to VML being underexploited. There are numerous projects within ONS that would benefit from a secure environment in which to process and analyse data, indicating that VML has the potential to become a valuable corporate asset provided it is correctly positioned and promoted.

Similarly it does not make sense that the funding of the VML should remain entirely the responsibility of the Economic, Labour and Social Analysis directorate. With the ESRC sponsored Secure Data Service taking over support for the majority of academic researchers it is likely that the main role for VML will be in supporting the development of official statistics with only a subsidiary role in supporting data analysis for research. Therefore the recommendation is that the VML should be funded centrally as a corporate service.

The real potential of VML in terms of cost savings, impact and maximum return to investment is as a service for the GSS as a whole. If the VML were to form the basis for a GSS wide archiving, development and analysis facility, not only would the costs of the service be shared across government, but bringing all government data sources together in one place has the potential for significant cost savings in data collection and respondent burden. With all data sources in one place there is more scope for matching and combining data to fill existing gaps, duplication of questions can be more easily identified, and planning for future resource investment can be carried out more effectively.

If these and other recommendations are acted upon, and the ONS is able to preserve the expertise developed by VML staff then the service will also be able to meet these ONS Aims and Strategic Objectives

- *To ensure IT supports business needs*
- *To build customers needs into our business plans*
- *To make the best use of our budgets*
- *To minimise the respondent burden (Whitestone 2010)*

To conclude in the VML the ONS has an internationally respected, relatively low cost²³ service that has been described by the OECD as 'one of the most innovative research efforts in the public sector across the 20 OECD countries'²⁴ (VML 2007). The service meets many of the ONS strategic and business aims; is a key infrastructure for core ONS projects; and provides a valuable link with academic experts. Discontinuing financial support for the VML is unlikely to be a cost saving for the ONS in the short or long term. The service is only in need of a small amount of refocusing to become a vital ONS asset.

²³ In comparison to similar services internationally

²⁴ In a letter from the OECD to Tony Clayton the previous head of EID, reported in VML annual report 2006/7

References

- (2007) Statistics and Registration Service Act 2007,
http://www.legislation.gov.uk/ukpga/2007/18/pdfs/ukpga_20070018_en.pdf
- Bloom, N., Sadun, R. & Van Reenen, J (2007) 'Americans Do I.T. Better: US Multinationals and the Productivity Miracle' *CEP Discussion Paper 788*, April.
<http://cep.lse.ac.uk/pubs/download/dp0788.pdf>
- Brandt, M. & Eilsberger, P. (2009) 'The ESSNET project – decentralized access to EU microdata sets' *UNECE Work Session on Statistical Data Confidentiality 2009*, Eurostat. (forthcoming)
<http://www.unece.org/stats/documents/ece/ces/ge.46/2009/wp.17.e.pdf>
- Brandt et al (2010) 'Guidelines for the checking of output based on microdata research' an *ESSNet report*. http://neon.vb.cbs.nl/casc/ESSnet/guidelines_on_outputchecking.pdf
- CO (2008) 'Data Handling Procedures in Government: Final Report' *Cabinet Office Report*
http://www.cesg.gov.uk/products_services/iatp/documents/data_handling_review.pdf
- Desai, T. & Ritchie, F. (2009) 'Effective Researcher Management', *UNECE Work Session on Statistical Data Confidentiality 2009*, Eurostat. (forthcoming)
<http://www.unece.org/stats/documents/ece/ces/ge.46/2009/wp.15.e.pdf>
- Dunnell, K (2008) 'Letter to Ian Diamond re Secure Data Service' 13.11.2008
- Elias, P. (2006) 'A National Strategy for Data Resources for Research in the Social Sciences', *Economic and Social Research Council*, Swindon (Report) http://www.uptap.net/wordpress/wp-content/National_Data_Strategy_Final_tcm6-16581.pdf
- Elias, P. (2009) 'A revised UK National Strategy for Data Resources for Social and Economic Research, 2009-2012 (report)' *Economic and Social Research Council*, Swindon (Report)
http://www.esrc.ac.uk/ESRCInfoCentre/Images/ESRC_NDS_publication_tcm6-34101.pdf
- Helpman, E. et al (2008) 'International Benchmarking Review of UK Economics', *Economic and Social Research Council*, Swindon (Report) http://www.esrc.ac.uk/images/IBR_economics_tcm8-4664.pdf
- ICO (2010) 'Data Sharing Code of Practice' *Information Commissioner's Office*, Consultation Paper
http://www.ico.gov.uk/about_us/consultations/~media/documents/library/Corporate/Research_and_reports/data_sharing_code_of_practice_consultation_paper.ashx
- Jackson, P. (2008) 'The Statistics Act and Research Access to Data' presentation at *ESDS Government meeting on changes in access to the government surveys and Labour Force Survey/Annual Population Survey user meeting* Thursday 17 January 2008, Royal Statistical Society, London
<http://www.ccsr.ac.uk/esds/events/2008-01-17/jackson.ppt>
- Marrano, M., Haskel, J. & Wallis, G. (2007) 'What Happened to the Knowledge Economy? ICT, Intangible Investment and Britain's Productivity Record Revisited' *Department of Economics, Queen Mary, University of London, Working Paper 603*, June.
<http://www.econ.qmul.ac.uk/papers/wp/WP603.HTM>

Martin, R., de Preux, L. & Wagner, U. (2009) 'The Impacts of the Climate Change Levy on Business: Evidence from Micro-data', *CEP Discussion Paper 917*, <http://cep.lse.ac.uk/pubs/download/dp0917.pdf>

ONS (2010a) 'Reducing the Administrative Burden on Business Resulting from ONS Surveys' *Office for National Statistics Compliance Plan 2010*, http://www.statistics.gov.uk/downloads/theme_other/ons-compliance-plan-2009-10v1.pdf

ONS (2010b) 'Taking Stock of Government Investment in Longitudinal Data: Key findings of a cross-government survey of investment in longitudinal data and analysis' *ONS Report for the Departmental Directors of Analysis Network (DDAN)*

Ritchie, F. (2002) 'MRP(03)03b-ext Business microdata access principles – draft proposals' *submission to ONS Statistical Policy Committee*.

Ritchie, F. (2004) 'Business Data Linking – Recent UK experience', *Austrian Journal of Statistics* v33:1-2 pp89-97.

Ritchie, F. (2008) 'Secure access to confidential microdata: four years of the Virtual Microdata Laboratory' in *Economic and Labour Market Review*; Office for National Statistics; May, pp 29-34 http://www.statistics.gov.uk/ELMR/05_08/downloads/ELMR_May08.pdf

Ritchie, F. (2011a) 'The VML Standard Operating Protocols – default procedures', *Office for National Statistics*.

Ritchie, F. (2011b) 'Provision of ONS data for analysis: Safe use, not safe data' *Office for National Statistics*.

Thomas, R. & Walport, M. (2008) 'Data Sharing Review Report' *Ministry of Justice Independent Review*, <http://www.justice.gov.uk/reviews/docs/data-sharing-review-report.pdf>

Van Reenen, J (2010) 'Letter in response to the ONS work programme consultation' 21.12.2010

VML (2007) 'Virtual Micro-data Laboratory Annual Report 2006/2007' *Office for National Statistics*

Whitestone, B (ed) (2010) 'ONS Business Plan 2010/11', *Office for National Statistics Publication*. <http://www.ons.gov.uk/about/our-organisation/corporate-documentation>

Annex A: Statistical Leadership - Organisations consulting MAUS

Statistics Canada
 DWP
 HMRC
 Mexico Statistical Office
 New Zealand Statistical Office
 Portuguese Statistical Office
 Secure Data Service
 Spanish Statistical Office
 OECD

Contributed to discussions in: Canada, China, Germany, New Zealand

Contributed to formal reviews in: Canada, New Zealand

Training programme adopted by: DCLG, ESS (European Statistical System), HMRC, Secure Data Service

Annex B: Glossary of Acronyms

BIS	Department for Business, Innovation and Skills
CO	Cabinet Office
DWP	Department for Work and Pensions
EID	Economic Interpretation Division
ELS	Economic, Labour and Social Analysis Directorate
ESRC	Economic and Social Research Council
ESSNet	European Statistical System Network
EU	European Union
GSI	Government Secure intranet
GSS	Government Statistical Service
HMRC	HM Revenue and Customs
LFS	Labour Force Survey
MAUS	Microdata Analysis and User Support
MSIP	Migration Statistics Improvement Programme
OGD	Other Government Department
ONS	Office for National Statistics
SDC	Statistical Disclosure Control
SDS	Secure Data Service
SRSA	Statistics and Registration Services Act 2007
UK	United Kingdom
UKDA	UK Data Archive
VML	Virtual Microdata Laboratory

Acknowledgements

My thanks to the ESRC and the ONS for funding the Knowledge Exchange Fellowship which led to this report, I am also very grateful to the MAUS team at the ONS who made me feel very welcome, in particular I'd like to thank Thea Thomas for all her work, Felix Ritchie for making things interesting and Richard Welpton for a Welsh social life. I would also like to thank Lesley Lilley for her support. My employer the Centre for Economic Performance at the London School of Economics generously agreed to do without data support for the period of my secondment for which I am also very grateful.