Strengthening Data Security: an Holistic Approach

Sharon Bolton & Matthew Woollard[1]

Abstract
In the light of heightened concern around data security, this paper highlights some of the measures that can be used to develop and strengthen security in data archiving. The paper includes discussion of the different approaches that can be taken towards the construction of firm and resilient data and information security policies within the social science data archiving communities. While international standards can provide theoretical guidelines for the construction of such a policy, procedures need to be informed by more practical considerations. Attention is drawn to the necessity of following a holistic approach to data security, which includes the education of data creators in the reduction of disclosure risk, the integration of robust and appropriate data processing, handling and management procedures, the value of emerging technological solutions, the training of data users in data security, and the importance of management control, as well as the need to be informed by emerging government security and digital preservation standards.

Keywords
Data security; data archiving; information security; data handling; user training.

New legislation and the data security 'climate'
During 2007, a series of high-profile data losses by UK government and associated organisations took place, involving reputedly 37 million items of personal data[2] In a climate of increasing concern over data security and identity theft, legislation in the form of the Statistics and Registration Services Act 2007 (SRSA) was at that time already making its way through the UK Parliament, its provisions intended to become effective from 1 April 2008. As a result of media furore over the data losses, in November 2007, the Cabinet Office was instructed to review data handling within government departments and make recommendations for their improvement where needed. This review was to be conducted under the direction of Robert Hannigan, Head of Security, Intelligence and Resilience. An interim report[3] was produced in December 2007, followed by the final report[4] in November 2008. Together, the SRSA and Cabinet reports had a marked effect on how data are now handled across UK government departments and culminated in the 'Mandatory Minimum Measures' for data handling[5] The SRSA also established the UK Statistics Authority, and for the first time introduced criminal penalties for the unlawful disclosure of confidential information (i.e. data relating to, or identifying, a particular person or business held or disclosed by the Statistics Authority).

The UK Data Archive (UKDA) is the curator of a large collection of UK government social science research data, held for use by the academic community. It is also licensed as a legal Place of Deposit by The National Archives, allowing the UKDA to ingest and preserve public records. Therefore, while not strictly covered by the Statistics and Registration Services Act, by the nature of its business the UKDA is intimately concerned with data integrity and security. While the UKDA has developed and maintained robust security practices over the years since its inception, it was felt that the time was right to review practices in response to the challenge of new legislation.

During the course of its work, the UKDA acquires data and associated materials from data creators, conducts ingest processing to prepare those data for secondary use, and supports users once they have received the data. Therefore, a similar holistic approach to the audit and refinement of data security was taken, to ensure that all the UKDA's data acquisition, ingest, access and support activities are supported by coherent and consistent procedures that work at all stages of the data archiving life-cycle.

Data creators and security enhancement
Firstly, the initial part of the process was reviewed – the work that the UKDA undertakes with data creators – and an assessment was made of how the new UK legislation on data security may affect practice. The UKDA will advise data creators at all stages of their project: from the planning stage, throughout the data gathering process and after completion and deposit of the data. This helps to ensure respondent confidentiality while maintaining sufficient detail within the data to enable effective research. The work is wide in scope, as data are acquired from a range of sources including large, well-funded government organisations, established research centres, and independent

small-scale academic research projects.

The first tangible effect of the new UK legislation that the UKDA experienced was a marked tightening of physical data transfer process from government, including the increased use of file encryption, more secure methods of data delivery, use of courier services, etc. This was unsurprising given the recommendations of the Cabinet Office report. The UKDA accordingly streamlined and secured acquisition and internal data transfer procedures to accommodate these developments.

Beyond the security of data transfer, new developments in the nature of data released from the Office for National Statistics (ONS) also quickly became apparent. Since before 2006, the ONS Microdata Release Panel (MRP) have overseen the release of ONS datasets by providing advice and testing for statistical disclosure, and as a result of the SRSA, the Panel updated their policy for the release of data at certain access-controlled levels. The UKDA were primarily concerned with two 'levels' covered by this data access control strategy, which have been developed by negotiation between the ONS and the UKDA: the End User Licence and the Special Licence. Datasets are compiled by ONS to these respective levels of detail prior to deposit with the UKDA.

To be able to download and use data from the UKDA, each user must register an account with the Economic and Social Data Service (ESDS - for which the UKDA is a service provider), via the UKDA website, and agree to the terms of the End User Licence. This includes an undertaking that the user must preserve at all times the confidentiality of information pertaining to, and not to attempt to identify, individuals and/or households in the data collections, nor must they share data with others who are not registered users. In accordance with the terms of how the SRSA defines levels of information, ONS data made available for research access at End User Licence level must not reveal or have the potential to reveal the identity of an individual[6] However, it is also recognised that researchers may sometimes need access to more finely-detailed data to conduct their research, and for this reason the Special Licence was set up with ONS. To gain access to data held under a Special Licence (which may include data designated as 'personal information' under the SRSA, but must have had some degree of protection applied), the prospective user must apply via the UKDA to ONS for Approved Researcher status. The candidate must provide extensive details of their academic background, status, and prospective research, and prove that they are a 'fit and proper' person to receive Special Licence data.

There are significant differences between the End User Licence and Special Licence versions of a dataset. As a practical example, the lowest geographic level permitted for End User Licence data is generally Government Office Region (GOR, or Nomenclature of Territorial Units for Statistics (NUTS) level 1) and education or employment and other demographic data may be banded or aggregated. Special Licence data, by contrast, may include geographic data at a finer resolution, such as Unitary Authority and NUTS2 or NUTS3 geographies, and more detailed education, employment and demographic data. Of course, the advent of the Special Licence does bring an added administrative burden; separate holdings of Special Licence and End User Licence versions of a dataset require doubled ingest processing work to be undertaken at the UKDA, and the administration of Approved Researcher applications is resource-intensive for both the UKDA and ONS. However, these steps must be taken to ensure that in a climate of heightened concern around security, data of a sufficient level of detail remain potentially available to the research community while respondent identity is still protected. Further to this (whilst it lies outside the remit of this paper) the UKDA will launch the Secure Data Service (SDS) in late 2009, where more potentially disclosive data will be available to selected Approved Researchers, who will then undergo further levels of security and system training.

However, not all government data acquired by the UKDA originate from ONS. While government organisations primarily concerned with data have robust procedures in place, other departments, especially those who have only recently begun to release data to researchers, are still coming to grips with the provisions of the SRSA, and do sometimes need guidance. As a result of consultations with ONS over End User Licence and Special Licence data, the UKDA are uniquely placed to offer such guidance should potential confidentiality issues arise, and often do so. As well as offering practical advice on how to balance disclosure risk while maintaining useful detail within data, by means of data edits, access control or a mixture of both, the UKDA (alongside ESDS Government) are currently involved in the process of facilitating discussions between the ONS MRP and other government departments, so that all may benefit from the MRP's expertise. Additional negotiation and guidance is likely to occur as a result of the Secure Data Service with new licence agreements between the UKDA and the ONS and between the UKDA and researchers.

In the UKDA's experience, academic researchers also vary widely in data security expertise. Some are attached to large research centres with established data managers and procedures, and others may undertake small-scale team or solo projects. While the data UKDA receive from government are largely quantitative, the data generated by academic projects may be quantitative, qualitative or a mixture of both. A considerable amount of work is undertaken by the UKDA to educate and inform the academic community on all aspects of data management, including obtaining consent; maintaining confidentiality and security; holding well-publicised regular workshops

covering both quantitative and qualitative data are held around the UK; and providing advice to individuals and organisations at all stages of the research process. The data deposited at the UKDA as a result of unique academic projects may present unusual challenges, and any potential edits or levels of access control that may be needed to reduce the risk of disclosure are discussed with researchers prior to the commencement of full ingest processing. The recent *Managing and Sharing Data* [7] guide gives straightforward and plain English advice on security issues that affect researchers.

Internal data handling

The second part of the holistic approach the UKDA took to strengthen data security was to audit all aspects of 'in-house' data handling. Members of staff at the UKDA have a wide range of expertise and experience, and work on diverse data tasks. Therefore, it was essential as part of the audit process to examine and scrutinise internal procedures, both human and technological, for the handling and storage of dataset files and associated administrative materials. As a result of this, existing good practice was identified and additional methods developed. These were collated into a comprehensive set of data security procedures, which were then distributed to all UKDA staff. In addition to the requirement that all staff register as UKDA users and are thus bound by the same conditions of the End User Licence, a Confidentiality Agreement has been also introduced. This details the responsibilities of staff with regard to data confidentiality, and requires the signature of the individual staff member. The smooth introduction of such stringent security measures needs to be carefully handled, and staff professionalism, awareness and expertise must be respected and acknowledged. Explanations as to why security measures were to be strengthened were given to staff, and all background information on the SRSA and the Cabinet reports have been made available, including legal aspects regarding differential treatment of data at the End User Licence and Special Licence levels. Staff members were encouraged to ask questions and actively feed into discussions throughout the process, and where needed, training was provided. Feedback from staff at all levels has resulted in some excellent suggestions that have since been incorporated into the security procedures. At the same time, a UKDA Security Plan[8] has also been developed in conjunction with the planning of the Secure Data Service. This plan brings together all aspects of information security within a single document and is based on the ISO 27001 standard. While the Plan has been developed explicitly for the Secure Data Service, it has ramifications which extend across the whole of the UKDA's procedures. It is within physical and environmental security and operations management that the key changes are likely to impact, but there are also technical changes that will be necessary and will influence the way in which staff carry out their daily work, even if their work is entirely unrelated to the Secure Data Service. The Plan will also be revised to take account of the more recent Her Majesty's Government (HMG) Security Policy Framework[9] and the provisions made in that Framework will have to be applied to the UKDA's interactions with government departments as well as researchers.

While this paper does not cover technological solutions in detail, it must be emphasised that the UKDA servers and system architecture and associated infrastructure are maintained to broadly conform to ISO 27001/2, and are updated periodically in step with technological advances. Regular systems testing is undertaken to identify any vulnerabilities, and other practical measures include the inclusion of integrated checksums in the names of downloaded files to prevent SQL injection, and processes put in place to prevent cross-site scripting and OS Command Injections. These 'development' and technical solutions are are under continuous development, and there will also be additional changes before the commencement of the Secure Data Service. The Security Plan also dovetails with other UKDA Policies and Procedures including the established UKDA Preservation Policy,[10] which covers data authenticity/integrity arrangements.

All plans, policies and procedures need regular review and updating. All the external factors that influence internal UKDA procedures and policies will continue to develop, and the UKDA must keep pace with these developments to both maintain internal standards and promote external confidence in the work of the UKDA. This ongoing process of review and revision will remain collaborative; all staff will continue to be encouraged to reflect on issues relating to data security that affect their roles and to use that knowledge to improve working practice. The successful introduction and implementation of new and sometimes tiresome procedures related to security needs should have considerable staff support, and it is only by allowing staff to contribute to the process that this support can be effectively harnessed.

Data security: The user's perspective

The UKDA's work with data users comprises the third major area for data security review and development. As there are currently over 50,000 registered UKDA/ESDS users, providing data security advice and support to users is a considerable task. The UKDA alongside its ESDS partners, acts to represent the interests of data users and accordingly facilitates dialogue with data creators to ensure users are given access to sufficiently-detailed data to permit useful research to take place. However, this brings an associated responsibility to promote safe data practices and train users in effective data security. The UKDA have developed measures to facilitate this, including regular workshops and training, and a web-based guide to good practice on micro data handling and security [11], which has been available for some years and is regularly updated to take account of new developments. These measures will be

extended considerably for the Secure Data Service. It is the UKDA's opinion that mandatory training courses for users of this service will provide one of the main planks in the strategy of allowing researchers desk-top access to secure data. The combination of detailed training, implementation of secure technologies, strong penalties and conformance with relevant standards will provide the necessary checks and balances to provide the most viable user experience.

Training in best practices in data management leads to training for researchers about the risks of disclosure. For the Secure Data Service it will be imperative that researchers are aware of what constitutes disclosure, how to they can identify disclosure risks in their own analyses, and what the legal and practical consequences of disclosure are. Training users will reduce the risk of security breaches. As part of the audit of data security, further measures are currently under consideration to improve the range of tools to educate data users on robust data security practice.

However, an effective policy has to be in place to deal with any sanctions and breaches of data confidentiality by users, who are made aware of potential measures that can be taken against them or their host organisation in event of a breach. Again, the planning for the Secure Data Service has identified new risks which will be covered by a new multi-tier procedure.

Conclusion

To summarise, it must again be emphasized that the work undertaken by the UKDA to strengthen data security has of necessity taken a holistic approach. The three fronts on which data security work is of prime concern (data creators, internal practices, and data users) are all interdependent. Work with data creators, including government departments at the beginning of the process, inculcates strong internal practices and data security awareness at the UKDA, which in turn leads to the better safeguarding of data and excellent educational work with data users. This paper has shown how data security at the UKDA has been strengthened in response to a changing external environment, and how the work will continue to develop as that landscape changes further.

Notes

1. Dr. Sharon Bolton, Data Services Manager, UK Data Archive: email sharonb@essex.ac.uk.Dr. Matthew Woollard, Associate Director, Head of Digital Preservation and Systems, UK Data Archive: email matthew@essex.ac.uk

2 Harrison, D. (2008) 'Government's record year of data loss', Daily Telegraph, 7 January. Retrieved 15 May 2009, from http://www.telegraph.co.uk/news/newstopics/politics/1574687/Governments-record-year-of-data-loss.html

3 Cabinet Office (2007) Data Handling Procedures in Government: Interim Progress Report, Cabinet Office, December. Retrieved 15 May 2009, from http://www.cabinetoffice.gov.uk/media/65934/data_handling.pdf

4 Cabinet Office (2008) Data Handling Procedures in Government: Final Report, June (published November). Retrieved 15 May 2009, from http://www.cabinetoffice.gov.uk/media/65948/dhr080625.pdf

5 Cabinet Office (2008) Cross Government Actions: Mandatory Minimum Measures, retrieved 20 May 2009, from http://www.cabinetoffice.gov.uk/media/cabinetoffice/csia/assets/dhr/cross_gov080625.pdf

6 Abrahams, C. and Mahony, K. (2008) New Policy and Procedures Governing the Release of Microdata Derived from ONS Social Surveys, paper presented at the 13th GSS Methodology Conference, London, June 23rd.

7 UK Data Archive (2009) Managing and Sharing Data.

8 The UKDA Security Plan is currently an internal document only.

9 UK Cabinet Office (May 2009) HMG Security Policy Framework, v.2.0 , retrieved 20 May 2009, from http://www.cabinetoffice.gov.uk/media/207318/hmg_security_policy.pdf

10 UK Data Archive (2008) UK Data Archive Preservation Policy. Retrieved 15 May 2009, from http://www.data-archive.ac.uk/news/publications/UKDAPreservationPolicy0308.pdf

11 Economic and Social Data Service (2008) Guide to good practice: micro data handling and security . Retrieved 15 May 2009, from http://www.esds.ac.uk/news/publications/microDataHandlingandSecurity.pdf