Chapter 6

Governance of data collection, data linkages and access to data

Eighteen of nineteen countries reported that there are multiple authorities in custody of key national databases for population health and health care monitoring and research. Data custodians in all countries reported significant efforts to protect data. Nonetheless there is variation across data custodians in challenging areas of data security including practices to de-identify data to protect patient privacy so that the data can be used for monitoring and research; and provision of safe mechanisms so that researchers from other government ministries or from academia could access and use data. Some custodians manage risk by refusing data access while others would consider providing access to identifiable patient-level data. Several data custodians noted that fulfilling all the responsibilities associated with data protection is expensive and there are cost pressures. A few countries provide interesting examples of centralising the difficult tasks of linking data, de-identifying data and approving and supervising access to data that have the potential to standardise practices and be more efficient. The sharing of person-level data across borders for international comparisons is rarely reported and there were few examples of data linkages for multi-country comparative studies.

This chapter presents country experiences in the de-identification of data to protect the privacy of individuals; the development of secure facilities for access to data with high re-identification risk; project approval processes for data linkage projects; data security within public authorities holding data and when public authorities provide data to external researchers; and governance of multi-country studies involving personal health data.

Data custodians play a central role in balancing data privacy protection and use of data for monitoring and research as they are responsible for the collection, processing, analysis and dissemination of personal health data. In many countries, data custodians are also responsible for vetting project proposals for the use of data from government and private entities; maintaining a technical capacity to undertake data linkages; maintaining a technical capacity for data de-identification; providing data access modalities to internal and external researchers; and ensuring that through all of their activities the legal requirements for data security and data privacy protection are respected. Several countries noted that fulfilling all of these responsibilities is expensive and that pressure is mounting to trim expenditure. Further, expenses are particularly heavy in countries with decentralised administration of the health system. In these countries, data custodians at sub-national levels are also carrying out these responsibilities. Advancements in techniques to ensure privacy by design and the development of privacy-enhancing technologies may provide avenues to meet both health care data use and privacy protection needs.

Other actors also play important roles in the governance of data collection and use, from legislators who establish governing legal instruments, to privacy regulators who ensure legislations are respected, to, for some countries, delegated bodies who review and approve proposals related to the development and use of personal health data, including independent bodies responsible for the implementation of national electronic health record systems.

De-identification of data

The practice of de-identification of data is widely used across the countries participating in this study; however, there is considerable variation in the interpretation of what constitutes de-identified data that may be legally released from a data custodian to an external researcher. The following are a few examples of different views.

France has invested in methods for data de-identification. This includes a hashing algorithm that converts names to a numerical code that cannot be reversed. These codes are then used to build longitudinal health histories. Given, however, that it is sometimes necessary in a research study to go back and verify content within clinical records, France has developed a reversible hashing algorithm for patient names. Such a reversible code is used, for example, by the Institut National de Veille Sanitaire during the first year of data processing for HIV positive patients, as the patients may need to be contacted by clinicians. After one year, the reversible code is erased and only an irreversible code remains on the file.

In *Finland*, data is considered de-identified when the identity number has been encrypted and names have been removed. Researchers outside of the National Institute of Health and Welfare (NIHW) with approved projects receive data with encrypted identity numbers to conduct their analysis. In *Sweden*, data is de-identified by the National Board of

Health and Welfare by removing national identity numbers, names, addresses and full dates of birth. Files provided to analysts within government and outside of government contain a study number that has been assigned in place of the identity number as well as some personal information on sex, age and home community. In *Denmark*, the National Board of Health data is de-identified by removing names and exact addresses. The national Central Person Register number, however, will remain on the analytical file. This number reveals the sex and birth date of the person.

In *Australia*, the Australian Institute of Health and Welfare (AIHW) considers data to be de-identified when direct identifying variables have been removed, such as names and exact addresses. There are, however, efforts underway to further reduce re-identification risk, where necessary, by introducing additional data processing, such as rolling up response categories for sensitive variables. The de-identification rules used for linked data will depend on the requirements of the custodians of the data supplied to the AIHW and in some cases the outcomes of consultation with the researcher.

In *Korea,* personal information, such as the Resident Registration Numbers and names of people, and information on individual corporations and organisations, such as health institutions, is strictly protected. In cases where researchers are approved to conduct a project involving data linkage, alternate keys are provided that cannot be used to identify individuals or organisations are provided. Further, access to data is only provided in a designated place within the Health Insurance Review Agency (HIRA).

In the *United Kingdom*, the NHS National Services Scotland (NSS) has identified certain fields within personal health data as sensitive (names, health numbers, full birth dates, and addresses). The NSS disclosure review protocol is applied to any personal health data to be disseminated outside of the NSS, which can result in suppression or treatment of variables that may pose a re-identification risk. For approved projects, researchers generally receive from the NSS a file where identifiers have been removed and where the health number has been replaced with a study number. The NHS Information Centre for Health and Social Care reported a similar process (see Chapter 2).

In *Canada*, the Canadian Institute for Health Information (CIHI) accepts encrypted and unencrypted health insurance numbers on administrative health databases that are transferred to CIHI from provinces and territories. When provinces and territories submit encrypted health numbers to CIHI, the encryption algorithm is maintained. When provinces or territories submit unencrypted health numbers, CIHI will encrypt the numbers using an established algorithm. CIHI's privacy disclosure procedures for the provision of de-identified data to third-parties, such as researchers, require the use of project-specific identification numbers instead of encrypted health insurance numbers. Where approved, an external third-party researcher may link databases using project-specific identification numbers.

In the *United States*, the National Centre for Health Statistics considers that data is de-identified when the risk of potentially re-identifying persons within the data has been reduced. This includes removal of identifiers, such as names, exact addresses, full dates and any identifying numbers and also a careful review of possible combinations of remaining sensitive variables within the data file that may indirectly lead to the disclosure of the identity of a person. Individual-level data that has been de-identified to this standard can be made publicly available and can be disseminated over the Internet to the public. For example, the linkage of population survey data to death data has been released

as a public-use micro data file. Often, however, the level of detail that is required for an approved research project would create a re-identification risk that is too high for the NCHS to release the data to the researcher. Instead, the NCHS has created a network of secure research data centres that researchers with approved projects must use. Similarly, in *Singapore*, researchers with approved projects may access de-identified data in the Ministry of Health's secured data lab (see below).

Secure facilities for access to data with a high re-identification risk

Custodians of personal health data in the United States, Canada and Singapore have created secure facilities where approved researchers may access de-identified personal health data that is deemed to have a higher than acceptable risk of potentially re-identifying individuals. This step has enabled the custodians to minimise the risk of misuse of the data.

The *United States* National Centre for Health Statistics has created a network of secure Research Data Centres (RDCs) across the United States in partnership with the US Census Bureau. In the RDCs, government and non-government researchers with approved projects access personal data necessary for their project and conduct all of their research. Only aggregated results may exit the facility after they have been reviewed by an NCHS staff member for any risks to data confidentiality. The NCHS has also introduced a new secure remote data access option for researchers, so that it is no longer necessary for all work to take place within the physical locations of the RDCs. Instead, researchers access a secure system called Andre from their own office. Through Andre they may submit programmes to analyse the data and receive the output. The Andre system has an automated process for checking for and preventing misuse of the data. Further, an NCHS staff member checks one-quarter of the data submissions and any detected misuse would terminate the researcher's access to the system.

In *Canada*, Statistics Canada also maintains a network of secure Research Data Centres across the Canadian provinces with similar features to the US RDCs (Statistics Canada, 2011). Researchers with approved projects may only have access to de-identified data with a high re-identification risk within the RDCs. Canada does not yet have a remote data access option, but is beginning to pilot options that may enable this type of access in the future.

In *Australia*, researchers with approval to undertake a project involving personal health data considered to be "high risk" would be able to access the linked de-identified data through a secure data linkage environment called SURE (Secure Unified Research Environment) that is offered through the Australian Population Health Research Network (PHRN). PHRN has received national government funding and is helping to advance data linkage infrastructure at the state and national levels. SURE is a new remote-access computing environment accessed via the internet. Access requires a username, password and authentication token. Key strokes on the local computer are transmitted to the SURE computer. Files cannot be transferred between studies or between the study and the local computing environment. Researchers can look at records on screen to resolve issues with their analysis, but they cannot print the screen or download any data. All outputs of their results are checked for confidentiality.

In *Singapore,* the Ministry of Health has also established a secure data laboratory that has been available for the past year. The ministry was concerned with the risk of

re-identification resulting from data involving the ministry's administrative databases. All approved research by government and non-government researchers involving access to de-identified data must take place within the lab. Only aggregated results that have been vetted by a ministry staff member may exit the secure lab.

In the *United Kingdom*, Universities and the Scotland NHS have launched a new initiative, the Scottish Health Informatics Programme (SHIP), that aims to eventually provide researchers with remote access to de-identified data in a secure manner so that it can be accessed at a distance from the data custodian and in a manner where the researchers may use advanced statistical techniques (Scottish Health Informatics Programme, 2011). SHIP also aims to ensure that data is shared across multiple custodians for linkage-based research and will be consulting with the public to define a transparent and publicly acceptable approach to the governance of this research.

Project approval process for data linkages

Across countries where research proposals for data linkages from external researchers may be approved, proposals must specify the data elements that are absolutely needed for their research and must justify the purpose and merits of their project in terms of the public interest.

In *Singapore*, all projects internal to the ministry and those from other governmental and non-governmental researchers involving linkages to the Ministry of Health's databases would have to be approved internally to ensure that linkage and access is legally permissible. Researchers with approved project would access the linked data in the ministry's secure data lab.

In *Korea*, a deliberation committee of the Health Insurance Review and Assessment Service (HIRA) approves data linkage projects on a project-by-project basis in accordance with the requirements of the Privacy Protection Act. Government and non-government researchers external to HIRA, such as non-profit academic researchers or researchers within public-good institutions may apply to the HIRA deliberation committee for access to de-identified personal health data including linked data that HIRA has in its custody.

In *Belgium*, the Privacy Commission approves data linkage projects. Approved projects that are part of the work programme of the Belgian Cancer Registry can have linkages undertaken by the Cancer Registry. Approved projects proposed by government or non-government researchers external to the Cancer Registry would be undertaken by the E-health platform. The platform would then provide de-identified data to the researcher for analysis.

Each registry in *Finland* has one person within it who is qualified to review project proposals for data linkages for scientific merit. If a researcher wishes to have data linked across several registries, the project proposal must be approved by the reviewer of each registry to proceed. All projects receiving approval are then sent to the national Data Protection Authority and the authority has 30 days on which to comment. The same approval process is followed for researchers within government and those outside of government. In *Sweden*, project proposals from within and from outside of government are reviewed and approved by the National Board of Health and Welfare. In *Denmark*, the Danish Data Protection Agency approves proposals for data linkage projects from within and outside of government. Researchers with approved projects then make a request for data linkage to the National Board of Health and Welfare.

In the *United Kingdom*, the UK Data Protection Act provides the legal framework wherein a national approach to decision making about data linkage projects could be developed. The Health and Social Care Act 2008 created the National Information and Governance Board (NIGB). NIGB was a national decision-making body for projects undertaken in the public sector or in the private sector where the consent of the data subjects was not obtained and where the use of the data was not authorised in law. The NIGB Ethics and Confidentiality Committee acted as a national research ethics approval body for all data custodians responsible for health and social care data. Thus, projects initiated by the public or private sector were reviewed for their conformity with the law; and the relative balance between research that is in the public's interest and the respect of privacy principles was weighed. For data files outside of the domain of health and social care, or for regions outside of NIGB jurisdiction (Scotland), the Caldecott Guardian would act as the approval body. Each custodian of personal data is required by law to have a Caldecott Guardian which is a senior official entrusted to protect data privacy and who is responsible for evaluating and approving projects requiring access to and use of personal data. The Health and Social Care Act 2012 transferred the functions delegated to the NIGB Ethics and Confidentiality Committee to the Health Research Authority (HRA) as of 1 April, 2013. The HRA is to protect and promote the interests of patients and the public in health research. It will streamline the current approval system and improve the efficiency and robustness of decisions about research projects (Department of Health, 2011). Changes to the constitution of the National Health Service have been proposed to offer patients a fuller explanation of their rights under existing law and NHS commitments with respect to data. A review of health information governance with an independent panel of experts will make recommendations on the balance between sharing personal information and protecting individuals' confidentiality (Department of Health, 2013).

In *France*, la Commission Nationale de l'Informatique et des Libertés (CNIL) is the French national data protection authority authorised by the *Loi Informatique et Libertés* (Data Protection Act). CNIL is an independent administrative authority that authorises, on a case by case basis, whether projects requiring access to identifiable personal health data will be approved. CNIL has a committee of experts in medicine and research which may advise on the scientific merit of proposed projects. Consideration for approval includes the legality of the request and the legitimacy of the researcher, including whether the researcher is affiliated with a credible organisation and the security measures that will be put into place to protect the data. Further, researchers requesting access to national health insurance data must also demonstrate that they have some authority that permits access to the data, such as an authorising legislation or professional membership. In addition to CNIL approval, non-government researchers must also be approved by the Comité du secret of the National Council for Statistical Information (CNIS).

Australia reports a complex system of project approval steps for researchers within and outside of government. An accredited Integrating Authority (such as the Australian Institute of Health and Welfare or the Australian Bureau of Statistics) will link data for an approved project that includes personal administrative data held by Australian Government agencies as well as state-level authorities. To be approved, however, the researcher has to demonstrate to the accredited Integrating Authority that approval has been secured from all of the data custodians and relevant Human Research Ethics Committees. For example, a linkage of the national cancer database to Pharmaceutical Benefits Scheme (PBS) records would require the approval of all eight cancer registries,

state-level approval from one or two authorities, as well as the federal Department of Health and Ageing (the data custodians of the PBS). Human research ethics approval may also be required by each data custodian, particularly in cases where the researcher requires patient consent requirements to be waived. In total, up to 20 separate approvals may be needed. Hospitalisations data must also be requested at the state level for national data linkage projects and obtaining essential approvals would be similarly onerous. An accredited Integrating Authority may return a de-identified linked data file to a researcher for use if the linkage does not involve a "high risk" database; if all of the data custodians involved agree; and if the researcher has the consent of study participants or has received a waiver to patient consent from a Human Research Ethics Committee. Research with "high risk" databases may only occur using a secure on-site data laboratory or within the secure remote data access facility called SURE (see previous section). To access SURE, researchers must also complete the SURE application process, sign an agreement of use and successfully complete SURE user training.

In *Canada*, the Canadian Institute for Health Information will review applications from internal and external researchers in both the public and private sectors for access to personal health data. In all cases, the researcher must apply for access and must justify each of the databases and data elements within the databases that would be required for the project. The researcher must sign a non-disclosure/confidentiality agreement that binds them to data security and confidentiality protection requirements and must commit to a time limit within which the data must be destroyed. CIHI can audit the researchers and researchers are aware of this possibility. Only de-identified data would be provided to the researcher.

In the *United States*, researchers wishing access to de-identified data that carries a re-identification risk must apply to the National Centre for Health Statistics (NCHS) for access to the data. NCHS management, and for some requests its internal review board, will review the research proposal and, if approved, the researcher will be provided access to the data within a secure Research Data Centre or within NCHS headquarters. It is also possible for a researcher to request a customised data linkage and the same process for approval would apply.

Most study participants indicated that commercially motivated research involving requests for access to identifiable data would fail to be determined to be for the public good and be rejected. In *Finland*, requests by commercial interests are ruled out. This is an issue because there is a law requiring pharmaceutical companies to conduct drug safety studies. To comply with that law, these companies would need to analyse personal health data from public registries. There are two solutions available now. The company could be identified as a scientific research centre, but this would be quite rare. Second, the company could hire a university researcher as a third party who could be approved to access data and report only aggregated statistical results back to the company. *Sweden* also does not rule out requests from commercial interests and reports a concern that it is difficult to sometimes ascertain if a research request for access to personal health data from a pharmaceutical company is really in the public's interest or if it is for commercial purposes and should be denied. To address this concern, Sweden is considering introducing new legislation to make clearer the conditions for access to personal data for research and analysis. In the *United Kingdom*, requests for data linkage by commercial interests are not ruled out, however they are more likely to fail to make a case that the request is in the public interest and therefore to not be approved. The Clinical Practice Research Datalink

will provide services to researchers in pharmaceutical or medical devices industries, subject to approval and compliance with the law (Clinical Practice Research Datalink, 2013). Written agreements bind researchers to conditions of access to data that include not using data to profile practitioners nor to evaluate advertising campaigns or the effectiveness of sales forces.

The specific case of researchers requesting linkage of their own data cohort

External researchers often request to have a cohort of data they have collected linked to public health databases. A very common occurrence is a request for the linkage of a clinical database or a database of clinical trial participants to subsequent hospitalisations, diseases and death. Such linkages will provide very important information about the effectiveness and safety of treatments and clinical care. At the same time, such linkages pose additional risk to data protection because the researchers involved have a strong ability to re-identify data within a de-identified database.

Virtually all countries that will provide researchers with access to linked data will consider such a request for approval. In all cases, however, the requesting researcher must be able to demonstrate that they had collected the data with the informed consent of the data subjects or had legal authorisation. In *Italy*, however, there are no routine or standardised procedures for a researcher to request a linkage of their own cohort of data to governmental databases and it seems that this type of project is impossible. In *Belgium and France*, the Privacy Commission renders a decision on all project proposals and would hear the proposal.

In *Australia*, the AIHW may agree to conduct a data linkage project involving a researcher's own cohort of data if all data custodians involved have approved the linkage and if a waiver of the need for consent has been provided by all human research ethics committees involved. Further, in Australia, deceased persons are not within the scope of the national Privacy Act (1988) and the AIHW ethics committee has approved that death data may be linked without consent under certain conditions. For similar reasons, the *United States* also reports that death data may be linked to a researcher's cohort without consent.

The *Switzerland* Statistical Office notes that such requests can be costly and that the time required to execute the requests is recovered from the researchers. This practice was also noted by *Denmark*. *Finland* noted that the National Institute for Health and Welfare is trying to keep costs low for external researchers but is under financial pressure. Some countries noted the challenge of charging for data that is a public good, even if the cost of custom data linkages is high. *Australia* reports that it intends to charge user fees for the SURE remote data access facility.

Data security within public authorities

In all of the countries participating in this study, data security and the protection of data confidentiality is given considerable attention. It was common for countries to report that their institution's existence or its ability to continue its programme of work would be placed at risk by any serious breach in data security. The elements of data security identified are accompanied by examples provided by country experts during the telephone

interviews. The next section discusses the specific case of data security for de-identified data provided to external researchers.

1. *Require employees to sign a non-disclosure or data confidentiality protection agreement.*

   The Australian Institute of Health and Welfare; the Belgian Cancer Registry; the Canadian Institute for Health Information; and the UK NHS NSS Scotland reported a requirement for new employees to sign a document that they will protect data confidentiality. The United States NCHS and the UK NHS Scotland reported an annual requirement for all employees to sign a document that they will protect data confidentiality.

2. *Provide staff with a written manual or a website describing their responsibilities for data confidentiality protection and security.*

   The US NCHS and the Australian AIHW have a staff manual on data confidentiality protection requirements. Data security and privacy guidelines are communicated to all employees of Korea HIRA using the internal network homepage. At the UK NHS NSS Scotland, standards for data protection and security are described in a document that employees must sign annually.

3. *Institute levels of approved access to data for staff.*

   At the Registerstele Krebsregister Schleswig-Holstein (Institute for Cancer Epidemiology) in Schlewig-Holstein, Germany; the Danish National Board of Health; the Finland National Institute for Health and Welfare (NIHW); the Australian Institute of Health and Welfare (AIHW); and the Canadian Institute for Health Information (CIHI); among others, individuals must be approved for access to data and only can see data relevant for their project requirements or job requirement. Some may have access to identifiable data, some to de-identified data and some have no data access at all. There are finer levels of approved access to data among employees of the Belgian Cancer Registry. Some employees may not see identifiable data; some may see identifiable data but only one record at a time and only to resolve data quality problems; and a small number of employees who work with physicians to receive data transfers and address quality issues may see identifiable data.

4. *Restrict data analysts from access to identifiable data.*

   At the Registerstele Krebsregister Schleswig-Holstein (Institute for Cancer Epidemiology) in Schlewig-Holstein, Germany, data analysts are never given access to personal identifiers and cannot access the computer system used by staff that process data. At the Swedish National Board of Health and Welfare, there is a specific statistical unit, the registry unit, which is permitted access to data containing identifying numbers. This unit cleans and processes the data and conducts data linkages for approved projects and de-identifies the data. Board analysts with permission to access files, see only de-identified data and never have access to the identified data. At the Singapore Ministry of Health, researchers and officers who perform analysis of de-identified linked datasets are restricted from accessing the identifiable constituent databases to minimise inadvertent re-identification and exposure of linked data records.

5. *Track and monitor staff access to data.*

   At the Singapore Ministry of Health, staff analysing linked data must do so from within a secure data lab. The use of the data within the lab is monitored and if there was any inappropriate handling of the data, it would be possible to identify the researchers

involved. Employees of the Belgian Cancer Registry with access to identifiable data must have their access logged. At the Swedish National Board of Health and Welfare, a security officer tracks which employees have been granted access to data. The National Board of Health in Denmark monitors who has access to registries and monitors and keeps logged how people with access are using the registry data on a 24/7 basis. The same protection and oversight applies to all national institutions in Denmark. The UK NHS Information Centre regularly reviews access logs to ensure that employees are still using the files that they are approved to access. A similar monitoring has also been introduced at the Swiss Federal Statistical Office.

6. *Provide training for new staff.*

   Staffs of the Canadian Institute for Health Information, the Belgian Cancer Registry and the Swedish National Board of Health and Welfare are trained in data security and confidentiality requirements when they are first hired. New employees of the National Board of Health in Denmark and the National Institute of Health and Welfare in Finland are trained in the use of data and data security by experienced colleagues.

7. *Provide refresher training for existing staff.*

   The US NCHS employees receive training on data security and confidentiality annually. Further, there are posters put up around the offices reminding staff about data confidentiality protection and security. The Belgian Cancer Registry provides training on global procedures regularly, including data security. The UK NHS Information Centre requires employees to take online training each year in data protection and then to pass a test. Every two months, employees of Korea's HIRA undergo data security and privacy training to ensure strict adherence to guidelines. The Canadian Institute for Health Information requires employees to complete mandatory on-line training annually on data privacy and security and to renew their pledge to protect data confidentiality. Ad hoc mandatory training may be required for some employees throughout the remainder of the year. The UK NHS NSS Scotland has on-line training in data security that is scenario based.

8. *Provide training for external researchers.*

   The Finland NIHW provides university-based researchers with a half-day or full-day training course on the NIHW databases, where part of the training is about data protection. The US NCHS requires researchers with approved access to a Research Data Centre to take training on data security and confidentiality annually. All researchers approved to undertake research requiring access to the SURE remote access facility in *Australia* will be required to successfully complete training on data confidentiality and security before accessing SURE.

9. *Require external researchers accessing data to become designated employees of the data custodian in order to place them under the same legal requirements and penalties as a regular staff member.*

   In the United States, contractors working for the NCHS who will touch data and external researchers approved to access de-identified data in the NCHS Research Data Centres must become designated employees of the NCHS. As a result, they are under the same legal obligations and penalties as staff of the NCHS to protect the confidentiality of the data they are working with.

10. *Secure buildings and offices*.

The German Institute for Cancer Epidemiology, where analysis of cancer registry data takes place at a national level, has strong physical security including doors that cannot be opened from the outside without a key. There is a clean desk requirement for staff engaged in data entry where no record can be left out at the end of the day. Records to be destroyed are stored in a separate container that cannot be easily accessed and a truck with a shredder comes monthly to security dispose of these materials. The Swedish National Board of Health and Welfare stores data in a building that is locked and secure. At the Finland NIHW, individuals may only share an office with another staff member who has approved access to the same data. Within the Australian AIHW, the Data Integration Services Centre (Data Linkage Centre) is physically separated from the other offices of the AIHW and only authorised personnel may enter the Centre.

11. *Secure transfers of identifiable data*.

In Sweden, data flows into the National Board of Health and Welfare are encrypted and sent in by mail. In Switzerland, the Federal Statistical Office uses secure servers to transfer data, for data storage and for access to data. In Finland, data flows into the National Institute for Health and Welfare (NIHW) take place using a secure electronic transfer. The UK NHS Information Centre uses a secure web transfer system similar to the older FPT protocol for data flows into and out from the Centre and protects the security of the system with a firewall.

12. *Secure computer systems for the storage of identifiable data*.

At the Singapore Ministry of Health, only data custodians and authorised data management staff are allowed access to identifiable personal health databases, meeting internal government standards for data security. The computer system used to process the identifiable data and conduct data linkages is completely separated from the computer system for analysis of de-identified data. The analysis of de-identified data takes place on standalone and isolated computers. In Switzerland, IT security requirements are under a specific federal department (IT) and all federal data is centrally stored and protected. Physical displacement of data is avoided. In Sweden, identifiable databases of the National Board of Health and Welfare are not stored on computers that are connected to a network, which protects the data from unauthorised access. At the Australian AIHW, the Data Integration Services Centre (Data Linkage Centre) has its own computer servers that cannot be accessed by staff outside of the Centre.

13. *Implement whole-of-government regulations or reporting up requirements on data security protection*.

The United States has federal regulations on data security that federal agencies must follow. The US NCHS must report to the government each year on its data security, and on any IT system changes that have occurred. The IT security is accredited every three years by the Centre for Disease Control. All federal agencies in the United States would have a similar oversight and monitoring of their IT security. Korea's HIRA has internal guidelines on the protection of data security and confidentiality including specific guidelines related to data linkage. Under the requirements of the new Personal Information Protection Act, the National Intelligence Service has issued guidelines on data security to government ministries including HIRA. HIRA will report annually to both the internal HIRA auditor and to the National Intelligence Service on its data security. The Belgian Cancer Registry has privacy and information security policies and a data

security plan required under the legislation authorising the registry. This plan is updated every three years. Elements of the security plan include how and when access to data is permitted; including levels of access to personal health data. The NHS NSS Scotland data security respects British Standards for Information Management and Data Sharing and NHS Scotland standards. In Singapore, there are guidelines within government for data protection.

14. *Institute third party or external data security audits.*

At the Belgian Cancer Registry, there are security audits by an independent organisation that will attempt to attack the security of the registry. The registry has received a high rating by the independent organisation for the results of its most recent security audit. In Korea, the Ministry of Health and Welfare, Ministry of Public Administration and Security and the National Intelligence Service have the authority to conduct privacy and security audits of HIRA. In Denmark, the Danish Data Protection Agency annually audits the National Board of Health to ensure that the handling of the databases meets legislative requirements. The Danish National Audit Office, which ensures that all national agencies comply with all relevant legislation may also audit the Board, or may rely on the results of the Data Protection Agency audit. In Australia, the data security environment of the Australian Institute of Health and Welfare (AIHW) was audited by a third party as a result of the AIHW's application to become a national integrating authority (national data linkage centre) for high risk linkage projects. Develop protocols in the event of a data security breach.

The United Kingdom NHS NSS Scotland and the NHS Information Centre have reporting systems that are used in the event of a suspected data security breach. In Korea, HIRA has a code to follow in the event of a data security breach.

15. *Institute legal penalties for deliberate breaches of data security.*

Within the US NCHS and Korea's HIRA, penalties for breaches of data security by employees include fines and imprisonment. Legal prosecution is also reported by the Danish National Board of Health as a consequence of a deliberate breach by an employee.

Data security when researchers receive data from public authorities

Data security is highest among data custodians requiring external researchers to access de-identified personal health data within a secure facility that is controlled by the data custodian or a third party. This practice was noted in the United States, Australia, Singapore and Canada. As discussed earlier, many data custodians provide approved researchers with access to de-identified data. Below are several examples of how data security is approached in this situation.

In *Finland*, when a researcher applies to access data, their application must demonstrate how their institution or university respects data protection requirements. Data is provided to the researcher on a compact disk that has been encrypted and the encryption key is provided to the researcher in a separate communication. Only identified and approved individuals who have been named may access the data.

In *Denmark*, the project approval will describe to the researchers the retention period of the file and will bind the researcher to not linking the data to any other databases and to not disclosing the data to a third party. The data protection authority in Denmark is then responsible for follow-up with the researchers to ensure compliance and data security

audits take place. Non-compliance is a legal violation and subject to penalties. At the data destruction date, the researcher will be given the option to de-identify the data, if they would like to retain the data for a longer period.

In the *United Kingdom*, the NSS Scotland indicates that the researcher is scrutinised during the approval process. A researcher who is a registered professional risks losing their profession as a result of a deliberate breach and, consequently, would be more likely to be approved. A researcher working within a recognised institution where data protection and data security are known to be high would also be more likely to be approved. Researchers sign their application that binds them to data security; to data confidentiality protection (including following rules for vetting any tables intended for publication); and to not share the data they have received with a third party. The NHS Information Centre for Health and Social Care indicated that there have been cases where linked data was given to a trusted third party for analysis, so that the risk of re-identification could be reduced.

In *Switzerland*, when data files are provided to an external researcher, their contract with the Federal Statistical Office binds them to protect the data and to follow the guidelines they are given. They are warned that they will be required to destroy the data if there is any infringement of these requirements. In practice, researchers want to be able to continue to collaborate with the Statistical Office and will follow the requirements. There is no audit of external researchers but there is tracking of their external publications to ensure that their use of the data is consistent with the agreed-upon purpose of their study.

In *Germany*, academic researchers can access de-identified personal health data for research. The provision of de-identified data for research is part of the laws that authorise cancer registries. While names will never appear on analysis files, some identifiers may be approved to remain on an analysis file, such as date and place of birth, if there is a justification for their inclusion in the research proposal. The decision to retain these identifiers will depend on the potential re-identification risk. Where re-identification risk may be high, solutions can include limiting the geographic variables to a higher level of geography or to retain only the month or year of birth.

Multi-country projects

Multi-country projects pose new challenges for data protection, as the data custodians involved typically have no legal recourse to exert any penalties for misuse of data by a foreign entity. Multi-country projects are difficult for research teams to implement, as the data protection requirements of each participating data custodian must be respected. Nonetheless, multi-country studies can provide a rich source of new information for the benefit of the public's health and the management of health systems and there are good examples of successful work.

The data protection legislations in some European countries make clear that it is possible to share identifiable data with other countries in the European Union. Noting this feature as part of national data protection legislations were *Denmark's* National Board of Health; *France's* Agence des Systèmes d'Information Partagés en santé and the *United Kingdom's* NHS NSS Scotland.

The *United Kingdom* NHS NSS Scotland indicated that under the UK Data Protection Act, it is not acceptable to share de-identified individual data outside of the European Union unless it can be demonstrated that the receiving country has the same standards for data protection as the United Kingdom. Some non-EU countries have been certified as

having equivalent standards and, for them, the process is the same as for an EU country. For a country not on the list, the two options for data access are a review of the country's legislation and an application for certification; or the provision of a fully de-identified data set, where there would be a very low risk of re-identification of individuals.

A similar process was reported by *France*. Under French law, the data protection authority (CNIL) may approve a project involving the sharing of personal health data with another EU country, as all EU members have established similar protections for data security and protection of privacy. If a project was to involve a non-EU country, the non-EU country would have to demonstrate that it has legislation that provides similar protection. For example, CNIL approved a project that involved sharing data with a researcher in the United States. Under a safe harbour agreement that was negotiated between the United States and the European Union, a contract was established that confirms that US laws (national and state) provide similar data privacy and security protection to those of EU countries.

Denmark's National Board of Health has contributed de-identified individual data to multi-country studies with other Scandinavian countries and has provided aggregate study results to multi-country studies led by many other countries including France, the United Kingdom and Germany. Similarly, the *Finland* National Institute for Health and Welfare has participated in multi-country studies based on data linkages. The *Belgium* Cancer Registry may contribute de-identified individual-level data to a multi-country study if the Office of Data Protection grants permission.

The *United States* National Centre for Health Statistics can provide a foreign researcher with access to de-identified individual-level data in two ways. In the first, the foreign researcher has equal access to public-use micro data files as does any domestic person. These files have been fully de-identified to result in a very low risk of re-identification of individuals. In the second, foreign researchers may submit a proposal to access data within the NCHS secure research data centres.

Australian researchers have participated in parallel studies where an Australian researcher received approval for the Australian data linkage and returned aggregated analysis to the multi-country study team. It is also possible for a foreign researcher to be approved to analyse *Australian* personal health data if the researcher is based in Australia and follows all of the same approval processes as any Australian national. The Australian Institute of Health and Welfare also recognises the legitimate need to share identifiable data across borders, particularly with New Zealand. There was a previous request to the AIHW from a New Zealand researcher for the linkage of a cohort of New Zealand military personnel to death records in Australia, as many had re-located to Australia and may have died there. The research ethics committee of the AIHW is reluctant to approve this request until they can be certain that legal penalties for any misuse of data by the foreign researcher can be applied.

There is an EU-funded project, EuroREACH, where representatives from participating countries in Europe and outside of Europe with experience in conducing national data linkage studies are working together to develop a website. The website would support researchers within and outside of government in the launch of multi-country health services research based on data linkages. It will draw on best-practice country examples in establishing comprehensive systems of performance measurement in European countries, and in granting research access to patient-level data for the study of health services. It will

also report on the person-level databases within countries that could support analysis and research and the steps required to produce population-based linked data sets and use them for multi-national health research projects (EuroREACH, 2011).

Bibliography

Churches, T. (2002), "The Use of Probabilistic Record Linkage, Public Key Cryptography and Trusted Third Parties to Improve the Protection of Personal Privacy and Confidentiality in Disease Registers and Tissue Banks", Symposium on Health Data Linkage Proceedings, Australian Government Department of Health and Ageing, Sydney, pp. 57-61.

Department of Health (2011), "Creation of the Health Research Authority", *www.dh.gov.uk/health/2011/12/creation-hra*, accessed 19 February 2013.

Department of Health (2013), "Information to Share or Not to Share? Information Governance Review", *http://caldicott2.dh.gov.uk/*, accessed 19 February 2013.

El Emam, K. (2008), "De-identifying Health Data for Secondary Use: A Framework", CHEO Research Institute, Ottawa.

El Emam, K., E. Jonker and A Fineberg (2011), "The Case for De-Identifying Personal Health Information", Social Science and Research Networks, *www.ssrn.com/absract=1744038*.

EuroREACH (2011), *www.euroreach.net*, accessed 25 October 2011.

Fraser, A. (2003) "Privacy and the Secondary use of Data in Health Research in Scotland", *Journal of Health Services Research and Policy*, Vol. 8, Suppl. 1, pp. 12-16.

Kalra, D., R. Gertz, P. Singleton and H.M. Inskip (2006), "Confidentiality of Personal Health Information Used for Research", *British Medical Journal,* Vol. 333, No. 7560, pp. 196-198.

Karmel, R., P. Anderson, D. Gibson et al. (2010), "Empirical Aspects of Record Linkage Across Multiple Data Sets Using Statistical Linkage Keys: The Experience of the PIAC Cohort Study", *BMC Health Services Research,* Vol. 10, p. 41.

Kelman, C.W., A.J. Bass and C.D.J. Holman (2002), "Research Use of Linked Health Data – A Best Practice Protocol", *Australian and New Zealand Journal of Public Health*, Vol. 26, No. 3, pp. 251-255.

Scottish Health Informatics Programme (2011), *www.scot-ship.ac.uk*, accessed 25 October 2011.

Statistics Canada (2011), "Research Data Centres", *www.statcan.gc.ca/rdc-cdr/index-eng.htm*, accessed 25 October 2011.