OECD Health Policy Studies

Health Data Governance

PRIVACY, MONITORING AND RESEARCH

OECD

OECD Health Policy Studies

Health Data Governance

PRIVACY, MONITORING AND RESEARCH

OECD
BETTER POLICIES FOR BETTER LIVES

This work is published under the responsibility of the Secretary-General of the OECD. The opinions expressed and arguments employed herein do not necessarily reflect the official views of OECD member countries.

This document and any map included herein are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

Please cite this publication as:
OECD (2015), *Health Data Governance: Privacy, Monitoring and Research*, OECD Health Policy Studies, OECD Publishing, Paris.
http://dx.doi.org/10.1787/9789264244566-en

ISBN 978-92-64-24455-9 (print)
ISBN 978-92-64-24456-6 (PDF)

Series: OECD Health Policy Studies
ISSN 2074-3181 (print)
ISSN 2074-319X (online)

The statistical data for Israel are supplied by and under the responsibility of the relevant Israeli authorities. The use of such data by the OECD is without prejudice to the status of the Golan Heights, East Jerusalem and Israeli settlements in the West Bank under the terms of international law.

Photo credits: Cover © vladvm50/Fotolia.com; © VectorShots/Fotolia.com; © iStock.com/geopaul; © Zern Liew/Shutterstock.com; Alexander Lukin/Shutterstock.com; © iStockphoto.com/GodfriedEdelman.

Corrigenda to OECD publications may be found on line at: *www.oecd.org/about/publishing/corrigenda.htm*.

Foreword

Health data collected by national governments that can be linked and shared are a valuable resource that can be used safely to improve the health outcomes of patients and the quality and performance of the health care systems that serve them. Data allowing a comprehensive view of health care services permit uncovering medical errors, adverse drug reactions, fraud, adherence to clinical guidelines, effective treatments, optimal care paths and optimal responders to treatment.

Health Ministry leadership is necessary to ensure that delivering the data to manage this important sector is at the forefront of government policy and action. Previous OECD work has found a high variability across OECD countries in data availability and use to concerns about and uncertainty about how to protect patient's rights to privacy and to preserve the security of health data when data are shared, linked and analysed.

This study supports OECD countries in developing privacy-protective uses of personal health data by examining current data availability, uses and governance practices; and identifying key data governance mechanisms that maximise benefits to patients and to societies and minimise risks to patients' privacy and to public trust and confidence in health care providers and governments.

International collaboration in this dynamic area is essential for information about best practices and lessons learned in health data governance to circulate widely; and to support movement toward common best practices so that multi-country statistical and research projects are feasible.

Acknowledgements

This OECD study was undertaken by the OECD HCQI (Health Care Quality Indicators) Expert Group as part of the 2013/14 programme of work of the OECD Health Committee. The OECD Working Party on Security and Privacy in a Digital Economy (SPDE) provided input to the study. The authors would like to acknowledge the representatives from the countries who make up the HCQI Expert Group and the SPDE Working Party who gave generously of their time to provide input to this study; and to the experts from participating countries that provided responses to surveys and participated in case study interviews (see Annexes B and C).

Particular thanks and recognition is extended to the members of the HCQI Advisory Panel of Experts on Health Information Infrastructure (APHII) for their contribution to all aspects of this study from its design and conduct, to analysis and reporting, to the development of tools and identification of key data governance mechanisms. Members of the APHII are recognised in Annex A of this report.

The contributions of the staff of the OECD Secretariat, in particular Jillian Oderkirk, Niek Klazinga, Anna Irvin Sigal and Duniya Dedeyn are also acknowledged. The authors would also like to acknowledge Stefano Scarpetta, Mark Pearson and Francesca Colombo for their support and directions provided on the study.

Table of contents

Tables

Figures

Acronyms and abbreviations

AHRQ	Agency for Healthcare Research and Quality (United States)
APHII	Advisory Panel of Experts on Health Information Infrastructure
CIHI	Institute for Health Information (Canada)
CPCSSN	Primary Care Sentinel Surveillance Network (Canada)
DPA	Data Protection Authority
ECHO	European Collaboration for Healthcare Optimisation
EEA	European Economic Area
EHR	Electronic Health Record
EMR	Electronic Medical Record
FSO	Federal Statistical Office (Switzerland)
GP	General practitioner
HCQI	Health Care Quality Indicators
HIPAA	Health Insurance Portability and Accountability Act (United States)
HIPC	Health Information Privacy Code
HIRA	Health Insurance Review and Assessment Service (Korea)
HMO	Health Maintenance Organisation
HSCIC	Health and Social Care Information Centre (England)
ICES	Institute for Clinical Evaluative Sciences (Canada)
IHIS	Institute of Health Information and Statistics (Czech Republic)
NBHW	National Board of Health and Welfare (Sweden)
NCHS	National Centre for Health Statistics (United States)

NHS	National Health Service (England)
PHC	Primary Health Care Organisation
PIPEDA	Personal Information Protection and Electronic Documents Act (Canada)
PROM	Patient-Reported Outcomes
RDC	Research Data Centre
RIVM	National Institute for Public Health and the Environment (Netherlands)
SAIL	Secure Anonymised Information Linkage (England)
STI	Sexually transmitted infections
SSI	Statens Serum Institute (Denmark)
THL	National Institute for Health and Welfare (Finland)
TTP	Trusted Third Party

Glossary

Cancer registry: A cancer registry is a type of patient registry defined by patients having a diagnosis of cancer. It is an organised system that uses observational study methods to collect uniform data (clinical and other) to evaluate specified outcomes for this population that serves a predetermined scientific, clinical, or policy purpose. The registry database is the file (or files) derived from the registry (ARHQ, *Registries for Evaluating Patient Outcomes: A User Guide*, 2007).

Cancer registry dataset: This dataset typically includes variables such as age, sex, location, date of diagnosis, method of diagnosis, site of neoplasm, type of neoplasm, stage and treatment.

Cardiovascular disease (CVD) registry: A cardiovascular disease registry is a type of patient registry defined by patients having a diagnosis of one or more types of cardiovascular disease, such as heart disease, acute myocardial infarction or stroke. It is an organised system that uses observational study methods to collect uniform data (clinical and other) to evaluate specified outcomes for this population that serves a predetermined scientific, clinical, or policy purpose(s). The registry database is the file (or files) derived from the registry (ARHQ, *Registries for Evaluating Patient Outcomes: A User Guide*, 2007).

CVD registry dataset: A CVD registry dataset may refer to only some cardiovascular disease conditions or to some procedures. For example, the European Society of Cardiology has developed a set of registries to assess cardiovascular risk factors, epidemiology and prevention measures; to monitor the application of clinical practice guidelines (heart failure, atrial fibrillation general, implantable cardioverter-defibrillation); and to assess the impact of interventional procedures and imaging techniques (atrial fibrillation ablation, transcatheter valve treatment) (ESC, www.escardio.org). Variables in such registries may include age, sex, risk factors, dates of diagnosis and treatment, method of diagnosis, procedures and treatment details and outcomes.

Clear data: Clear data refers to data that have been de-identified by the removal or pseudonymisation of direct identifiers but where dataset values remain original values and have not been perturbed or obscured by data masking techniques. Such data can carry a higher risk of re-identification and also a higher utility for statistics and research.

Clinical terminology classification system: Standard sets of terms, names and codes to be used for health care coding. For example, the WHO ICD (International Classification of Diseases) is often used for diagnosis coding; the WHO ATC (Anatomical Therapeutical Chemical Classification System) is often used for coding medicines; and SNOMED-CT (Systemised Nomenclature of Medicine – Clinical Terms) provides a broad set of standardised clinical terms for software applications and is increasingly used in electronic clinical records.

Confidentiality: Confidentiality relates to disclosure or nondisclosure of information. Historically a duty to honour confidentiality has arisen with respect to information disclosed in the context of a confidential relationship, such as that between an individual and his or her physician, attorney, or priest. In such relationships, the confidante is under an obligation not to disclose the information learned in the course of the relationship. Now the law applies such duties to some holders of information who do not have a confidential relationship to a patient. The importance of confidentiality to the medical profession is reflected in the physician's "Oath of Hippocrates".

Data confidentiality: Data confidentiality is a property of data, usually resulting from legislative measures, which prevents it from unauthorised disclosure.

Data masking: Data masking describes a set of techniques used to de-identify personal data by perturbing data values to reduce the likelihood that dataset records could be re-identified. Examples of data masking techniques include supressing variables or variable values, grouping variable values, restricting the range of variable values, swapping variable values among dataset records, rounding variable values or otherwise distorting variable values in a random manner.

Data protection: Data protection refers to the set of privacy-motivated laws, policies and procedures that aim to minimise intrusion into individuals' privacy caused by the collection, storage and dissemination of personal data.

Dataset record: A dataset record is a row of data in a dataset table consisting of a single value from each column of data in the table. The data in the columns of the dataset are all of the same type of data, such as birth date or address, whereas the rows represent a given instance, such as a single patient or person or a group of patients or persons.

De-identified data: This is data which do not identify an individual directly, and which cannot easily be used to determine identity. De-identification requires the removal of name and exact address; and can also involve the removal of any other detail or combination of details that might support identification.

Deterministic record linkage: In this approach, often referred to as exact matching, a unique identifier or set of identifiers is used to merge two or more sources of data. In health linkages, the identifier used is often a unique patient identifying number or UPI.

Diabetes registry: A diabetes registry is a type of patient registry defined by patients having a diagnosis of diabetes. It is an organised system that uses observational study methods to collect uniform data (clinical and other) to evaluate specified outcomes for this population that serves a predetermined scientific, clinical, or policy purpose(s). The registry database is the file (or files) derived from the registry (ARHQ, *Registries for Evaluating Patient Outcomes: A User Guide*, 2007).

Diabetes registry dataset: A diabetes registry dataset typically includes age, sex, risk factors, date of diagnosis, type of diabetes, lab tests, procedures, treatments and complications.

Direct identifier: A direct identifier is a means to identify a specific individual and can include their name, full address or unique patient identifying number (health insurance number, social security number).

Electronic Clinical Record: For this OECD study, an electronic clinical record includes clinical information about individual patients within electronic medical, patient or health records. See definitions of electronic health record and electronic medical record/electronic patient record.

Electronic Health Record: For this OECD study, an electronic health record (EHR) refers to the longitudinal electronic record of an individual patient that contains or virtually links records together from multiple Electronic Medical Records (EMRs) which can then be shared across health care settings (interoperable). It aims to contain a history of contact with the health care system for individual patients from multiple organisations that deliver care.

Electronic Medical Record/Electronic Patient Record: For this OECD study, an electronic medical record (EMR) or Electronic Patient Record (EPR) is a computerised medical record created in an organisation that delivers care, such as a hospital or physician's office, for patients of that organisation. EMR/EPR is provider or organisation centric and allows storage, retrieval and modification of clinical patient records.

Emergency care: Acute care of patients who present without prior appointment, either by their own means or by ambulance. Emergency care is usually found in a hospital [emergency department (ED), also known as accident & emergency (A&E), emergency room (ER), or casualty department] or other primary health care centre.

Emergency care dataset: This dataset will typically include information on the dates of attendance and discharge, reason for attendance, the diagnosis, treatments or procedures provided, medications at discharge, and discharge destination. It may also include information on waiting times and whether or not an ambulance was used.

Formal long-term care: Long-term care is the care for people needing support in many facets of living over a prolonged period of time. Formal long-term care can be provided in home, institutional or day-care settings, from public, not-for-profit and for-profit providers, with services varying from alarm systems to daily personal care.

Formal long-term care dataset: This dataset typically includes information on patient age and sex, main diagnosis, dates of care, care type and care provider. It may also contain information on the patient's functional health status and mental health status.

Health care coding: The process of assigning a standard code to a description of a clinical diagnosis, procedure or treatment using a standardised clinical terminology classification system. *(See definition of clinical terminology classification system).*

Health data: Health data usually consist of individual, personal health and other related information. The European Group on Ethics in Science and New Technologies (EGE), in the Opinion No 13 Ethical Issues of Health Care in Information Society defines “health data” as including “a wide range of information about an individual, which all touch upon an individual’s private life. A health biography could include not only basic medical data: a history of all medical diagnoses, diseases and medical interventions, medications prescribed, test results, including imaging, etc. but could also include more sensitive data: on mental health, relevant to family history, behavioural patterns, sexual life, social and economic factors, etc. and health care administrative data: admissions and discharge, data routine, operational data, insurance and financial transactional data, etc.

Hospitals: Hospitals comprise licensed establishments primarily engaged in providing medical, diagnostic, and treatment services that include physician, nursing, and other health services to in-patients and the specialised accommodation services required by in-patients. Hospitals may also provide out-patient services as a secondary activity.

Hospital in-patient dataset: This dataset will typically include information on the age and sex of in-patients, their dates of admission to hospital and discharge from hospital, their main diagnosis, the procedures administered to them and medications prescribed at discharge.

In-patient care: In-patient care refers to care for a patient who is formally admitted (or "hospitalised") to an institution for treatment and/or care and stays for a minimum of one night in the hospital or other institution providing in-patient care. In-patient care includes accommodation provided in combination with medical treatment when the latter is the predominant activity provided during the stay as an in-patient.

Mental hospital: Mental hospitals comprise licensed establishments primarily engaged in providing medical, diagnostic and treatment services that include physician, nursing and other health services to in-patients requiring care for mental health, psychiatric or substance-abuse related health conditions.

Mental hospital inpatient dataset: This dataset will typically include information on the age and sex of in-patients, their dates of admission to hospital and discharge from hospital, their main diagnosis, the procedures administered to them, and medications prescribed at discharge.

Mortality dataset: A census of all deaths by cause of death and demographic characteristics of the deceased within a defined population.

Network of health care organisations: A network of health care organisations provides a continuum of health care services. The network may provide integrated care under a parent holding company. Some networks have a Health Maintenance Organisation (HMO) component. Networks of health care organisations, such as Kaiser Permanente in the United States, offer a broad range of health care services and can conduct research where patient data are linked across the different health care facilities they operate.

Patient experiences survey dataset: This dataset contains the results of a survey to measure patient experience of health care services. Content domains can include accessibility of care, co-ordination of care, communication quality, adherence to clinical guidelines, and patient satisfaction. It includes both surveys of patients or of service users, including surveys of the general population.

Patient-reported outcomes (PROMs): Patient-reported outcomes (PROMs) are reports coming directly from patients about how they feel or function in relation to a health condition and its therapy without interpretation by health care professionals or anyone else. PROMs can relate to symptoms, signs, functional status, perceptions, or other aspects such as convenience and tolerability (*Cochrane Handbook for Systematic Reviews of Interventions*, 2008). Questionnaires are often used to collect PROMS both before and after a treatment is given.

Patient-reported outcomes (PROMs) dataset: PROMs may be collected from patients at the point of care or collected from patients via a telephone, mail or other survey. Thus PROMs data may exist as a stand-alone dataset or PROMs data may be included within other datasets, such as within hospital datasets, primary care datasets, patient survey datasets or population health survey datasets.

Population census or registry: A population census is the total process of collecting, compiling, evaluating, analysing and publishing or otherwise disseminating demographic, economic and social data pertaining, at a specified time, to all persons in a country or in a well delimited part of a country.

Population census or registry dataset: This dataset typically includes variables such as age, sex, location, household members, education, employment, income, ethnicity, and immigration status. When integrated with or linked to health data it provides a powerful means to understand differences in health and health outcomes within a population, such as socio-economic disparities in health outcomes or access to care.

Population health survey dataset: This dataset contains the results of a survey of the general population regarding health status and presence of diseases, socio-demographic characteristics and, in most cases, exposure to health-related risk factors.

Prescription medicines: Prescription medicines are medicines exclusively sold to customers with a medical voucher, irrespective of whether it is covered by public or private funding and include branded and generic products.

Prescription medicines dataset: This dataset will typically include information on prescription medicines dispensed over the counter (community pharmacies) by their name or by code, as well as the date of dispensing.

Primary care: Provision of continuing and comprehensive medical care to individuals and families in an ambulatory setting. It may be provided by general practitioners (or "family doctors") and their teams. The critical elements are a focus on the part of the providers on generalism rather than specialism; the provision of patient-centred rather than disease-centred, co-ordinated, and accessible services; and the integration of biomedical, psychological, and social dimensions of the presentation and management of presenting problems.

Primary care dataset: This dataset will typically include information on the age and sex of patients, the dates of visits, the main diagnosis, medications prescribed, and lab and imaging test results.

Privacy: Privacy is not being observed or disturbed by others. Privacy is a concept that applies to data subjects, while confidentiality is a concept that applies to data.

Probabilistic record linkage: In this approach, a set of possible matches among the data sources to be linked are identified. For example, identifying information such as names, dates of birth, and postal codes, may be used to assess potential matches. Then statistics are calculated to assign weights describing the likelihood the records match. A combined score represents the probability that the records refer to the same entity. Often there is one threshold above which a pair is considered a match, and another threshold below which it is considered not to be a match. This technique is used when an exact match between records across databases is not possible, or when data capture errors have caused deterministic matches to fail.

Pseudonymisation: This is a technique where identifying information about individuals, such as names, complete addresses and patient numbers are converted to a meaningless name or number in a consistent manner. The consistency of the application of the pseudonymisation algorithm permits record linkage among databases. The assignment of a pseudonym may be done it a way that permits it to be reversible or not.

Record linkage: Record linkage refers to a merging that brings together identifiable records from two or more sources of data with the object of consolidating facts concerning an individual or an event that are not available in any separate record (*Handbook of Vital Statistics Systems and Methods*, Vol. 1: Legal, Organizational and Technical Aspects, United Nations Studies in Methods, Glossary, Series F, No. 35, United Nations, New York, 1991.) An example would be linking patient records in a hospital database to any death records for the same persons in a mortality database in order to identify patients who died following treatment.

Regions/states: A region/state is an area or a division of a country. In some OECD countries, datasets unavailable at a national level are available at the region/state level. Most often, this is due to having delegated responsibility for health systems to these areas. Examples of sub-national areas with significant health data assets include: New South Wales Australia, Southern England, Piedmont Italy, and Ontario Canada.

Re-identification: Re-identification is attributing identifying variables to an individual's record within a de-identified dataset. Re-identification requires information about the individual obtained from personal knowledge or from data stored in other datasets about the same individual. For example, a person who is listed in a non-health dataset with their name and address included might be matched, with some probability, to a health dataset that has no names or addresses included. Using probabilistic record linkage, the two databases are linked to the same individual on the basis of similar variables available in both datasets. Examples of similar variables might be city, sex, age, marital status, diagnosis, etc.

Executive summary

OECD countries are ageing and increasing shares of our populations are living longer with multiple chronic and disabling conditions. This shift is placing pressure on limited health care resources. To meet this challenge, health system managers and policy makers are moving toward performance-based governance to improve care quality, co-ordination and efficiency. Performance-based governance requires timely and accurate patient data that span the continuum of care, including health outcomes and costs. Such data also support re-designing and evaluating new models of health care service delivery and contribute to the discovery and evaluation of new treatments.

While all countries are investing in health data infrastructure, there are significant cross-country differences in data availability and use, with some countries standing out with significant progress and innovative practices enabling privacy-protective data use; and others falling behind with insufficient data and restrictions that limit access to and use of data, even by government itself. Countries that develop a data governance framework that enables privacy-protective data use will not only have the information needed to promote quality, efficiency and performance in their health systems, they will become a more attractive centre for medical research and will have opportunities to build public-private partnerships.

To support OECD countries in improving data governance frameworks, health ministries and data privacy protection experts in OECD countries collaborated in 2013/14 to pursue this in-depth investigation to understand the current situation, to uncover and document practices, and to identify promising data governance mechanisms that enable privacy-protective monitoring and research. Advice and guidance on all aspects of this study were provided by a multi-disciplinary panel of experts.

Countries that have developed strong health data governance frameworks provided good examples of how data can be used safely to benefit society. Overall, among the 22 states participating in this study, the health information systems with the greatest data availability, maturity and use were found in Denmark, Finland, Iceland, Israel, Korea, New Zealand, Norway, Singapore, Sweden and the United Kingdom (Wales and Scotland).

After examining the current situation in OECD countries, data governance mechanisms were identified to maximise societal benefits and to minimise societal risks from uses of health data. These mechanisms build forward from existing efforts, such as the OECD Privacy Framework (OECD, 2013) and the European Data Protection Directive (95-46-EC), to begin to address an unmet need for an international consensus about effective practices in the protection of privacy in the use of personal health data, so that we may facilitate greater harmonisation of privacy-protective monitoring and research activities. The mechanisms should assist countries developing governance frameworks and engaging in legislative reforms, including those necessary as the result of the anticipated EU Data Protection Regulation.

The Advisory Panel of Experts on Health Information Infrastructure identified key data governance mechanisms supporting privacy-protective data use:

1. The health information system supports the monitoring and improvement of health care quality and system performance, as well as research innovations for better health care and outcomes.
2. The processing and the secondary use of data for public health, research and statistical purposes are permitted, subject to safeguards specified in the legislative framework for data protection.
3. The public are consulted upon and informed about the collection and processing of personal health data.
4. A certification/accreditation process for the processing of health data for research and statistics is implemented.
5. The project approval process is fair and transparent and decision making is supported by an independent, multidisciplinary project review body.
6. Best practices in data de-identification are applied to protect patient data privacy.
7. Best practices in data security and management are applied to reduce re-identification and breach risks.
8. Governance mechanisms are periodically reviewed at an international level to maximise societal benefits and minimise societal risks as new data sources and new technologies are introduced.

Each mechanism is the focus of a chapter of this report and the chapters conclude with the essential elements of the mechanism that could be included in a national framework to strengthen health information infrastructure. To support balanced decision making about the approval of projects involving the processing of personal health data, a *Risk-Benefit Evaluation Tool* is also provided (Chapter 6, Table 6.2).

International collaboration in this dynamic area is essential for information about best practices and lessons learned in health data governance to circulate widely; and to support movement toward common best practices so that multi-country statistical and research projects are feasible.

This study reveals several areas where international collaboration is needed, in particular to:

- support countries in developing the norms necessary for governments to certify or accredit data processors;
- develop guidance for the implementation of project approval bodies;
- ensure that there are sufficient agreed international standards for data coding and interoperability;
- support countries to evaluate which national legal frameworks for the protection of health information privacy provide adequate protections to facilitate multi-country statistical and research projects;
- review current practices in patient consent and in waivers to consent to reach a common understanding about mechanisms that are privacy protective;
- review developments in data security risks and threats and mechanisms to address them; and
- explore mechanisms to engage the public in discussion about data and its governance to ensure that there is good public awareness of health data, the benefits of its use, its protection, and the rights of data subjects.

Chapter 1

Introducing high-value, privacy-protective health information systems

This chapter introduces the data that are essential to improving health care and health system performance and why the processing of these data poses risks to the protection of the privacy of data subjects. It presents the conceptual framework and methodology and summarises the content of the report.

The statistical data for Israel are supplied by and under the responsibility of the relevant Israeli authorities. The use of such data by the OECD is without prejudice to the status of the Golan Heights, East Jerusalem and Israeli settlements in the West Bank under the terms of international law.

Highlights

While all countries are investing in health data infrastructure, there are significant cross-country differences in data availability and use, with some countries standing out with significant progress and innovative practices enabling privacy-protective data use; and others falling behind with insufficient data and restrictions that limit access to and use of data, even by government itself.

To support OECD countries in improving data governance frameworks, health ministries and data privacy protection experts in OECD countries collaborated in 2013/14 to pursue this in-depth investigation into health data governance including the development and use of personal health data in OECD countries and the legal frameworks, policies and practices that are in place to protect the privacy of data subjects when data are being processed and analysed.

The purpose of this investigation is to understand the current situation, to uncover and document practices, and to identify promising data governance practices that enable privacy-protective monitoring and research.

The overarching framework for this study is that decision making about potential statistical or research uses of personal health data should be taken after considering both societal risks from the data use and societal benefits from the data use. Optimal decision making about potential statistical and research uses of data can only be achieved if there is an overarching data governance framework in the country that has itself been optimised to minimise societal risks from data use and to maximise societal benefits from data use.

OECD countries are ageing and increasing shares of their populations are living longer with multiple chronic and disabling conditions. This health shift has important implications for how care is best organised and provided; where new treatment innovations can be expected; and future cost pressures on governments. To address the burden of chronic conditions, medicine must focus on preventing their on-set and controlling their progression. At the same time, health systems must focus on improvements in care quality and co-ordination; and efficient care delivery and on finding new ways to make systems more productive and sustainable.

The need to more actively manage health system outcomes will drive health systems toward greater use of clinical and administrative data to assess the comparative-effectiveness of therapies and services. These data will also be needed to support re-designing and evaluating new models of health care service delivery and to contribute to the discovery and evaluation of new treatments.

Better data will be needed to assess and compare the effectiveness of therapies and services provided to chronically ill patients. Better data will also be needed to support re-designing and evaluating new models of health care service delivery and to contribute to the discovery and evaluation of new treatments.

Health Ministry leadership is necessary to ensure that delivering the data to manage this important sector is at the forefront of government policy and action. Effective collaboration between health ministries, justice ministries and data privacy regulators is essential if governments are to evolve toward a situation where societal benefits from data use are maximised and risks to society from data use are minimised. At the same time, government needs clear and open channels to engage with stakeholders in the development and use of data, so that data governance frameworks and practices reflect societal values and priorities.

In 2010, health ministers called for improvement in national information infrastructures to support research and monitoring to enable national health care quality and system performance improvements; and to strengthen the ability of their countries to develop internationally comparable indicators of health care quality. The motivation for this call was to shift away from cost containment in the management of health care to the generation of evidence about the outcomes of care for performance-based governance.

The OECD Health Care Quality Indicators Expert Group began surveying countries in 2011 regarding the development of national health data assets and their use to improve health, health care quality and health system performance (OECD, 2013). We found that while all countries are investing in data infrastructure, there were significant cross-country differences in data availability and use, with some countries standing out with significant progress and innovative practices enabling privacy-protective data use, and others falling behind with insufficient data and restrictions that limit access to and use of data, even by government itself.

To support OECD countries in improving data governance frameworks, health ministries and data privacy protection experts in OECD countries collaborated in 2013/14 to pursue this in-depth investigation into health data governance including the development and use of personal health data in OECD countries and the legal frameworks, policies and practices that are in place to protect the privacy of data subjects when data are being processed and analysed. The purpose of this investigation is to understand the current situation, uncover and document practices, and identify promising data governance practices that enable privacy-protective monitoring and research.

This effort has been led by the OECD Health Care Quality Indicators Expert Group. Advice and guidance on all aspects of the study was provided by the Advisory Panel of Experts on Health Information Infrastructure (APHII). APHII is a multi-disciplinary panel of international experts with backgrounds in health policy, research, statistics, law, privacy regulation, and information technology and includes representatives from government, academia, industry and civil society (see Annex A). The OECD Working Party on Security and Privacy in a Digital Economy has provided input to the study.

This chapter introduces the data that are essential to improving health care and health system performance and why the processing of these data pose risks to the protection of the privacy of data subjects. The conceptual framework and methodology for this OECD study is presented and the content of this report is summarised.

Data are essential to improving health care and health system performance

Essential to health care quality and performance assessment is the ability to follow patients as they progress through the health care system from primary health care to speciality care to hospitalisations, long-term care, home care, hospice care and death. These data should also provide information about underlying patient characteristics, illnesses, medications, therapies, tests and images. This type of follow-up permits a comprehensive view of health care services provided and the health outcomes of those services; and permits uncovering medical errors, adverse drug reactions, fraud, adherence to clinical guidelines, effective treatments, optimal care paths and optimal responders to treatment.

Understanding pathways requires linking datasets at the patient level, as current health data are usually collected in silos. As a result, key datasets about elements of the health care pathway must have sufficient detail to enable valid and reliable dataset linkages. The development and use of data from electronic health records (EHRs) has the potential to enable a quantum leap in health care quality and performance assessment because such records can be brought together into an electronic health record system that captures patients' health care pathways and outcomes and, from which, data can be extracted.

Progress toward linking data and extracting data from clinical record systems, however, remains limited in several countries.

High-value data about health care pathways and outcomes support discovery and innovation

In the past, stratifying patients into groups that share common characteristics, such as age, sex, disease history, risk factors, medications, lab or image results, has been difficult. One of the potential uses of high-value data is to uncover how different clusters of patients with different backgrounds and characteristics respond to existing therapies. It is through developing this understanding that more specific guidance can be provided to carers regarding the best therapies to recommend. At the level of the health system, this opens an opportunity to not only reduce harms and safety concerns but also to improve overall system efficiency by getting the right care to the right patient the first time.

When high-value data can be united with genomic information, even greater differentiation of effectiveness of therapies may be uncovered. Combined with clinical data, genomic data enable stratifying patients at the molecular level and provides the possibility to identify profiles consistent with increased risk of disease and with greater or reduced responsiveness to different treatments. Genomics also enables discovering new therapies at a genetic level to address disease risk or combat diseases.

A further use of high-value data is to identify finely disaggregated patient groups to be invited to specific clinical trials and to track the progress of trial participants over the long run.

We are only at the beginning of understanding how new technologies for remote monitoring including medical devices and apps could contribute to understanding how dynamics in health conditions, health behaviours and exposures to environmental harms impact upon our health and the safety, effectiveness and efficiency of health care treatments. Developing this understanding would require linking or integrating monitoring data with data about care pathways and outcomes.

Societies without good data risk poor health care quality and lost innovation

There are significant risks to individuals and to societies when health information assets are not developed, or are unused or are very difficult to use. Societies lose the opportunity to monitor and report on their population’s health and the quality and safety of health care services. This elevates the risk of individuals experiencing inefficient, ineffective and even harmful health care. Societies also lose the opportunity for research and innovation to improve health and health care outcomes, which can improve well-being, productivity and the efficient use of public resources.

Health data use may put patients’ privacy at risk

Historically a duty to honour confidentiality has arisen with respect to information disclosed in the context of a confidential relationship, such as that between an individual and his or her physician, attorney, or priest. In such relationships, the confidante is under an obligation not to disclose the information learned in the course of the relationship. Now the law applies such duties to holders of information who do not have a confidential relationship to a patient but where the data held is detailed enough to identify the data subjects, either directly or indirectly.

Health data that can be linked to measure pathways and outcomes are often both personal and sensitive. It is personal because there is information that identifies individuals and it is sensitive because it is about aspects of individual's health and health care treatments and services that they have received. In many cases, the data are an outcome of the confidential relationship between patients and their health care providers. Both the sharing and the linkage of such data risk the protection of the privacy of the persons whose data are involved.

When data are shared they may be lost or stolen during the transfer process or the data recipient may not provide sufficient protection to keep the data confidential. When data are linked, the combined dataset provides more information about the data subjects than did the original unlinked datasets. Thus the resulting linked data could cause more harm to data subjects if it were lost, stolen or otherwise misused.

Potential harms to individuals that could result from the misuse of their personal health information can be severe and can include financial and psychosocial harms. Financial harms can result from discrimination in health insurance or employment. Psychosocial harms could include embarrassment, stigma and loss of reputation, resulting in isolation and stress. Disclosures of personal data can also increase individual's risk of experiencing identity theft. Less discussed, but of social relevance, is also the risk of loss of public confidence in government and its institutions that could result from misuses of individuals' personal health records, including a loss of confidence in the health care system.

Maximising societal benefits and minimising societal risks

The overarching framework for this study is that decision making about potential statistical or research uses of personal health data should be taken after considering both societal risks from the data use and societal benefits from the data use (Figure 1.1). If both dimensions are not evaluated, then decision making is likely to be sub-optimal for society.

Benefits that may arise from data uses include promoting individuals' rights to health through improved therapies and higher quality and more efficient health care services; producing research and evidence that responds to societal values regarding health and well-being, safe and effective health care, scientific discovery and innovation, and efficient, accessible, affordable and co-ordinated health care services; and producing positive economic outcomes for health system actors, governments and the economy through efficiency gains, returns to discovery and innovation and savings in data collection costs.

Risks that may arise from data uses include infringements upon individuals' rights to privacy; decisions and processes that fail to respond to societal values regarding privacy and data sharing; exposures of individuals to lost privacy and other harms, such as discrimination, social stratification leading to class disparities or and identity theft; and decisions and processes that weaken societal trust in health care providers and governments.

An important dimension of the framework for this investigation is data governance. Optimal decision making about potential statistical and research uses of data can only be achieved if there is an overarching data governance framework in the country that has been aligned to minimise societal risks and to maximise societal benefits from data uses.

Countries that have developed strong health data governance frameworks provided good examples of how data can be used safely to benefit society. After examining the current situation in OECD countries, the APHII has identified eight data governance mechanisms to maximise societal benefits and to minimise societal risks from the use of health data (Box 1.1). These mechanisms build forward from existing efforts, such as the OECD Privacy Framework (OECD, 2013) and the European Data Protection Directive (95-46-EC), to begin to address an unmet need for an international consensus about effective practices in the protection of privacy in the use of personal health data, so that we may facilitate greater harmonisation of privacy-protective monitoring and research activities. The mechanisms should assist countries developing governance frameworks and engaging in legislative reforms, including those necessary as the result of the anticipated EU *Data Protection Regulation*.

Box 1.1. The eight key data governance mechanisms

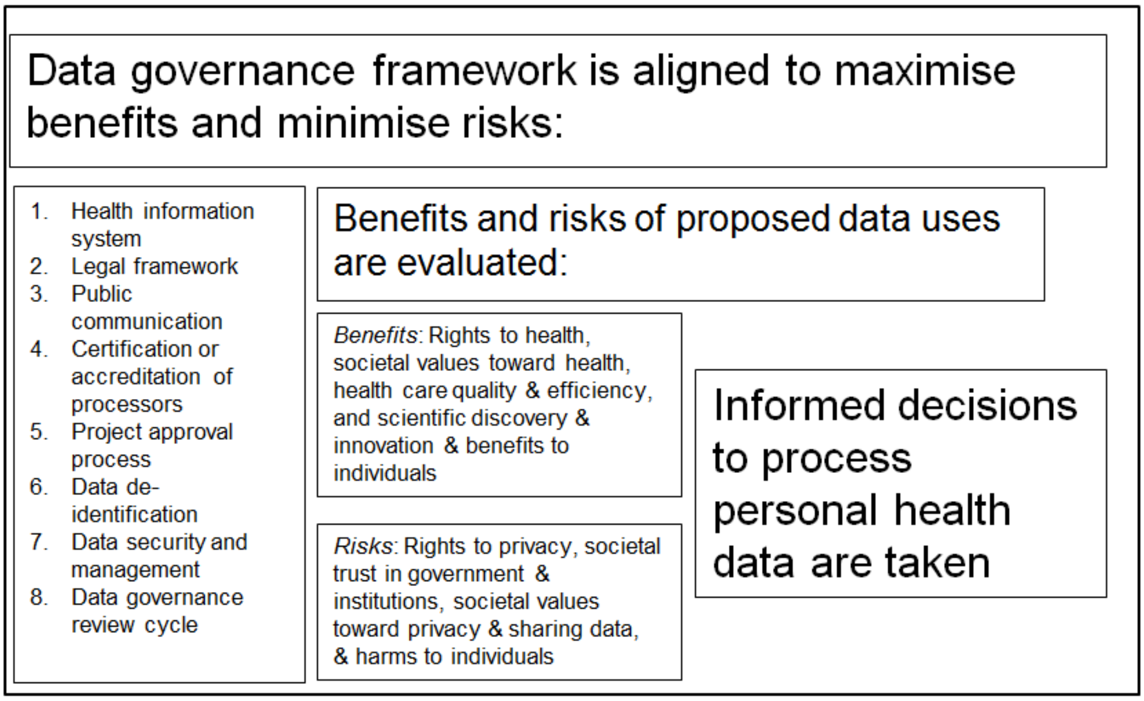
1. The health information system supports the monitoring and improvement of health care quality and system performance, as well as research innovations for better health care and outcomes.
2. The processing and the secondary use of data for public health, research and statistical purposes are permitted, subject to safeguards specified in the legislative framework for data protection.
3. The public are consulted upon and informed about the collection and processing of personal health data.
4. A certification/accreditation process for the processing of health data for research and statistics is implemented.
5. The project approval process is fair and transparent and decision making is supported by an independent, multidisciplinary project review body.
6. Best practices in data de-identification are applied to protect patient data privacy.
7. Best practices in data security and management are applied to reduce re-identification and breach risks.
8. Governance mechanisms are periodically reviewed at an international level to maximise societal benefits and minimise societal risks as new data sources and new technologies are introduced.

Data gathering and reporting

Through the Health Care Quality Indicator's Expert Group, OECD countries were invited to take part in an international survey in 2013 to describe the availability and use of personal health datasets within countries in order to monitor progress since 2011; to explore dimensions of national data governance; and to describe recent policy-relevant studies at the national and international levels requiring the processing of personal health data (Annex A). Twenty OECD countries completed a detailed questionnaire in 2013 regarding their data assets and how they are governed. Three members of the United Kingdom presented separate responses and were found to have important differences of interest to OECD countries. As a result they have been analysed separately in this report, bringing the total number of states included in this study to 22 (Canada, Czech Republic, Denmark, Finland, Iceland, Ireland, Israel, Italy, Japan, Korea, Netherlands, New Zealand, Norway, Singapore, Spain, Sweden, Switzerland, Turkey, the United States and the United Kingdom (England, Scotland and Wales).

These countries identified national experts in dimensions of data governance including legal frameworks for health information privacy protection, project approval processes, data de-identification, data security mechanisms, and data access mechanisms. Fifty-two experts were interviewed by telephone in 2013 and 2014 (Annex B).

Figure 1.1. Data use decisions should be taken by weighing societal benefits and risks within a data governance framework that maximises benefits and minimises risks


Data governance framework is aligned to maximise benefits and minimise risks:
1. Health information system
2. Legal framework
3. Public communication
4. Certification or accreditation of processors
5. Project approval process
6. Data de-identification
7. Data security and management
8. Data governance review cycle
Benefits and risks of proposed data uses are evaluated:
Benefits: Rights to health, societal values toward health, health care quality & efficiency, and scientific discovery & innovation & benefits to individuals
Risks: Rights to privacy, societal trust in government & institutions, societal values toward privacy & sharing data, & harms to individuals
Informed decisions to process personal health data are taken

The first draft report of the findings from the country survey and telephone interviews were shared with the Advisory Panel of Experts on Health Information Infrastructure (APHII). The APHII panel convened in person in Paris on 21 May 2014 to discuss together the elements of health data governance that help to maximise societal benefits while reducing societal risks and to develop a tool to evaluate risks and benefits of uses of personal health data.

In preparation for this meeting, APHII members were asked to consider the findings and respond to the APHII modified Delphi survey. In this survey technique, opinions of the APHII members were gathered regarding elements of a data governance framework that maximises societal benefits and minimises societal risks. It had two parts: Part A asked opinions on a taxonomy enabling countries to identify and evaluate the risks and benefits of proposed uses of personal health data within countries' existing data governance frameworks. The elements included were intended to support decision making about individual project proposals. Part B asked opinions about data governance mechanisms that could allow countries to optimise their data governance framework. The mechanisms included were intended to help countries to maximise societal benefits from personal health data while minimising societal risks.

Questionnaires were completed by APHII members independently and then collated and shared with all APHII members for consideration and review leading up to the meeting. Areas where views diverged were highlighted and were the focus of the discussion on 21 May. At the meeting, the APHII discussed revisions to the governance elements. Revised governance elements were then circulated to APHII members and, following written feedback from members, refined further and discussed at a web-conference of the APHII panel in September 2014.

APHII members participated in the drafting and revision of this report in 2014-15, in response to feedback from government officials responsible for health systems and data privacy protection and national experts in legal and operational aspects of health data governance.

Each of the eight key data governance mechanisms supporting privacy-protective data use that were developed by the APHII is the focus of a chapter of this report. Each chapter concludes with the essential elements of the mechanism that could be included in a national framework to strengthen health information infrastructure.

Chapter 2 focuses on the ***development of high-value data*** that describe patients' health care journeys and their outcomes over time. It describes the degree to which OECD countries have the underlying health information infrastructures to realise such data, as well as success stories from countries participating in national and international projects advancing high-value data to promote health and improve health care.

Chapter 3 presents the ***legislative frameworks*** in OECD countries related to the protection of health information privacy. It describes differences in data accessibility throughout the OECD that relate to legal frameworks and to their interpretation in practice; how countries handle situations where project-specific patient consent is not possible or practical; the degree to which identifiable and de-identified data can be shared and with whom, including commercial and foreign applicants for access to data; and challenges encountered in developing data sharing arrangements.

Chapter 4 describes the degree to which health information systems in OECD countries are ***open and are transparent*** regarding the data within the health information system, how it is being used, the safeguards surrounding its use and data subject's rights with respect to their own data.

Chapter 5 discusses the degree to which national health data ***processing is concentrated*** within countries, the advantages associated with a concentration of data processing and how processors can be accredited or certified to meet the country's highest standards for the protection of personal health data.

Chapter 6 describes ***the project approval processes*** of OECD countries including the involvement of research ethics boards, internal review boards and data privacy regulators in decision making and public transparency regarding the approval process. A ***risk-benefit evaluation tool*** is presented that countries could use to guide bodies evaluating and approving applications to process personal health data (Chapter 6, Table 6.2).

Chapter 7 provides examples of ***data de-identification processes*** and explains how there can be gaps between the goals of legislations and data de-identification practices; and the importance of considering the bigger picture of the potential benefits of the data use and the data security surrounding the data use when decisions about de-identification processes are taken.

Chapter 8 describes ***practices to protect the security and confidentiality of data*** within data processors and when data processors share data with third parties, such as other government departments and researchers. These include fundamental security elements such as controlled access to facilities and networks and training; public transparency about the protection of data; data sharing agreements, auditing for compliance and penalties for non-compliance; and alternatives to data sharing, including remote data access systems and secure research data centres.

Chapter 9 concludes the report by sharing views among OECD countries regarding the progress they have made toward privacy-protective data use and their ***outlook for the next five years.*** Elements of international collaboration that are essential to ensuring that governance mechanisms remain relevant over time and that support further progress toward privacy-protective data use are proposed.

References

CIHI – Canadian Institute for Health Information and Canada Health Infoway (2013), Summary and Recommendations for Moving Forward from *Better Information for Improved Health: A vision for Health System Use of Data in Canada*, Canadian Institute for Health Information and Canada Health Infoway, June.

El Emam, K. (2013), *Risky Business: Sharing Health Data while Protecting Privacy*, Trafford Publishing, United States.

OECD (2013), *Strengthening Health Information Infrastructure for Health Care Quality Governance: Good Practices, New Opportunities and Data Privacy Protection Challenges*, OECD Health Policy Studies, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264193505-en.

Chapter 2

High-value health data supporting health care management, policy and innovation

This chapter explores the strengths and weaknesses in OECD countries' health information systems and provides practical examples of strong health information systems and the benefits that have accrued to countries that have fostered them. Overall, the health information systems with the highest value in terms of data availability, maturity and use were found in Denmark, Finland, Iceland, Israel, Korea, New Zealand, Norway, Singapore, Sweden and the United Kingdom (Scotland and Wales).

Highlights

Ten countries reported having 70% or more of the key national health and care datasets necessary for understanding health care pathways and outcomes. The national personal health datasets reported by countries tend to have very high coverage of targeted populations; rely upon automatic data extraction from electronic clinical and administrative records; and include the use of standard codes for clinical terminology.

Thirteen countries are regularly linking data from at least four national datasets: hospital and mental hospital in-patients; cancer registry data and mortality data. Key reasons for approving these linkages include to develop health care quality and system performance indicators to measure care co-ordination, outcomes of care pathways, resource utilisation and costs, and compliance with national health care guidelines.

Twenty-seven examples of national projects involving the linkage of datasets or the extraction of data from electronic clinical record systems illustrate the potential for routinely collected data to improve our understanding of what works, for which patients, when and at what price. Examples of using such data to benchmark and compare the performance of countries internationally were also shared, with most projects having been conducted in Europe.

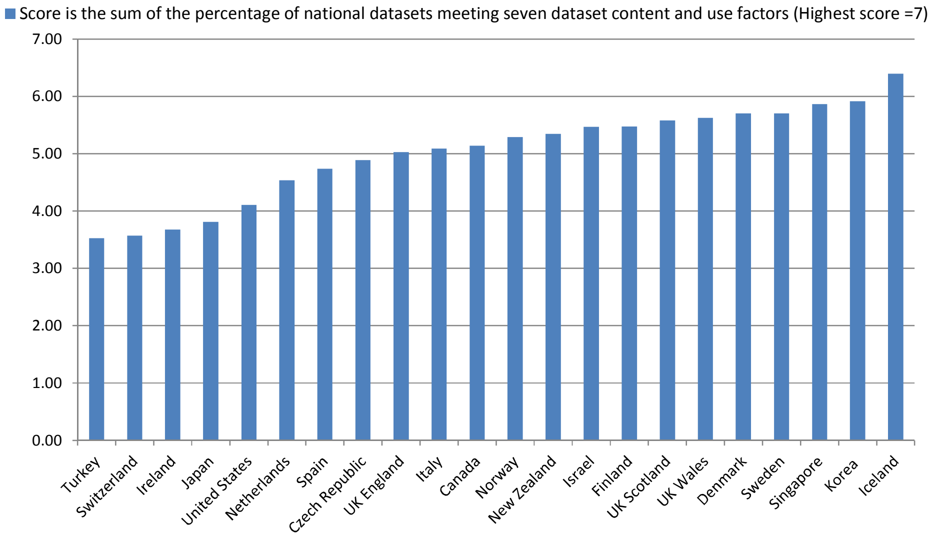
OECD countries are making tremendous investments in health data collections and information management systems. Nonetheless, many OECD countries have a poor track record in bringing these investments toward their full potential in terms of information value. Further, decisions taken about health data governance at the national level either impede or encourage data development and use and have a strengthening or a dampening effect on health data gathering and sharing across society including within sub-national governments and public bodies, the research community and within the private sector. Encouraging the uptake of the most efficient and effective frameworks and practices to enable the collection, storage and use of personal health data to improve population health and to improve the effectiveness, safety and patient-centeredness of health care systems remains a significant policy challenge in many OECD countries.

The OECD has been surveying countries about their health information assets and the use of these assets for statistics and research since 2011. This chapter provides an overview of the strengths and weaknesses in national information systems and provides examples of the benefits accruing to a limited set of pioneering countries.

Figure 2.1 provides a high-level summary of the strength of the health information systems across OECD countries in 2013. The figure presents a score for each country that is the sum of the proportion of the key national personal health datasets investigated that meet seven different development and use criteria measured in this study (see Table 2.1). These are the percentage of key datasets available with high population coverage; with automated data extraction from electronic record systems; with the same unique ID number; with standard codes for clinical terminology; and used for regular reporting of quality and performance and linked for regular quality and performance reporting.

Overall the strongest health information systems were found in Denmark, Finland, Iceland, Israel, New Zealand, Norway, Korea, Singapore, Sweden, and the United Kingdom (Wales and Scotland).[1] The most limited health information systems were found in Ireland, Japan, the Netherlands, Spain, Switzerland, Turkey and the United States.

Figure 2.1. Key health data availability, maturity and use



Source: Author’s own calculations from the results of this study.

Among these countries, there are important differences in performance for each of the seven factors investigated (Table 2.1). Each of the seven factors is described in the remainder of this chapter.

Table 2.1. Key national health dataset availability, maturity and use

	% of key national health datasets available[1]	% of health care datasets with coverage of 80% or more of the population	% of available health care datasets where data extracted automatically from electronic clinical or administrative records	% of available datasets sharing the same unique patient ID	% of available datasets where standard codes are used for clinical terminology	% of available datasets used to regularly report on health care quality or health system performance (published indicators)	% of available datasets regularly linked for research, statistics and/or monitoring (indicators)	Total
Canada	71%	60%	63%	50%	100%	100%	70%	5.14
Czech Rep.	50%	50%	60%	86%	100%	71%	71%	4.89
Denmark	86%	90%	78%	92%	100%	75%	50%	5.70
Finland	79%	90%	44%	100%	89%	55%	91%	5.47
Iceland	79%	90%	90%	100%	100%	91%	90%	6.39
Ireland	57%	38%	60%	0%	100%	88%	25%	3.67
Israel	64%	55%	67%	89%	83%	100%	89%	5.47
Italy	64%	70%	86%	44%	100%	100%	44%	5.09
Japan	71%	68%	86%	50%	86%	20%	0%	3.81
Korea	79%	80%	88%	91%	100%	82%	73%	5.92
Netherlands	57%	50%	83%	56%	86%	67%	56%	4.54
New Zealand	57%	59%	86%	75%	83%	100%	75%	5.35
Norway	100%	50%	79%	93%	100%	50%	57%	5.29
Singapore	71%	80%	88%	100%	88%	70%	90%	5.86
Spain	36%	30%	75%	67%	100%	100%	67%	4.74
Sweden	86%	90%	89%	83%	89%	67%	67%	5.70
Switzerland	50%	47%	80%	43%	80%	43%	14%	3.57
Turkey	100%	73%	100%	0%	80%	0%	0%	3.53
United States	64%	13%	57%	64%	86%	73%	55%	4.11
UK England	64%	28%	100%	78%	100%	44%	89%	5.03
UK Scotland	57%	61%	88%	100%	75%	100%	78%	5.58
UK Wales	64%	65%	100%	100%	100%	44%	89%	5.62

[1] Includes hospital in-patient data, mental hospital in-patient data, emergency health care data, primary care data, prescription medicines data, cancer registry data, diabetes registry data, cardiovascular disease registry data, formal long-term care data and mortality data.

Source: Authors own calculations based on the results of this study.

Key national health and health care datasets

In the 2013 OECD country survey, participating countries were asked about the availability, characteristics and uses of the following 14 key sources of national personal health data:

- hospital in-patient data,
- mental hospital in-patient data,
- emergency health care data,
- primary care data,
- prescription medicines data,
- cancer registry data,
- diabetes registry data,
- cardiovascular disease registry data,
- mortality data,
- formal long-term care data,
- patient-reported health outcomes data,
- patient experiences survey data,
- population health survey data and
- population census or registry data.

These datasets were identified because of their potential to provide high information value. In particular, they support both the potential to understand pathways of care and outcomes for all people and for groups of people with different characteristics. They are the essential building blocks for understanding what works? For whom? When? And why?

Pathways of care involve understanding health care from the patient's perspective which is the receipt of services, often from a set of providers and involving sets of therapies that have immediate and long-term consequences. Patients journey from diagnosis in primary care to specialist care to emergency rooms to hospital stays and to long-term care services and back and forth among these services and experience improvements and deteriorations in their health during the journey and afterward. The datasets included in this study cover the key health care services provided to patients: hospital in-patient services; community health services including primary health care, emergency health care and formal long-term care (such as nursing homes and home care services). The use of prescription medicines is a key part of the health care services offered to patients that are delivered in hospital, in other care settings and in the community to be used at home. They are both tremendously useful and highly risky products and understanding benefits and risks is essential to keeping patients healthy and safe. Thus these data are a key component of health care pathways and outcomes.

Health care paths must include the ultimate loss of health, which is death. Deaths occur inside and outside of hospitals and other health care settings. Key to keeping patients safe and in understanding the effectiveness of health care treatments is measuring patient survival following care and in the absence of necessary care.

Disease registries are a particular type of dataset where the data subjects are defined by having a particular diagnosis. While some registries offer only limited information to track disease incidence and prevalence; increasingly registries involve extraction of data from electronic medical records and database linkages to understand the natural history of disease, treatments and outcomes. In Finland, for example, morbidity registers are created by combining treatment data (medications, hospitalisations, and primary care) and data on causes of death. Disease registries are used to assess the effectiveness of treatments provided to patients as well as to identify patients for invitation to clinical trials. While many countries established cancer registries decades ago and have accepted them as essential to making progress in combatting this disease and improving patient outcomes; the same regard has not been paid to any other prevalent diseases facing OECD populations. Thus registries for two other prevalent health conditions were included in this project, diabetes and cardiovascular diseases, to demonstrate where there is progress in national monitoring of these conditions and, importantly, where there is none.

Contextual information is required to stratify patients to understand what works? For whom? When? And why? Population census or registry data provide detailed information about populations such as education, employment, income, housing, ethnic origin, language, immigration status and other details that are essential to evaluating whether or not health care services and health care outcomes are distributed equally or unequally. Such details are also essential for understanding linkages between socio-demographic characteristics of patients and their risks of developing disease or of dying. Similarly, population health surveys provide even more granular information about health risk behaviours and outcomes for understanding why access to care and outcomes of care may differ.

Patient experiences surveys, typically administered to selected patients experiencing different health care encounters or treatments, provide insight into the quality of services and may shed light on reasons why outcomes of care differ for different patient groups. Patient-reported outcomes data or PROMS are survey instruments where patients respond to questions designed to measure their functional health status before and then after the administration of health care therapies. PROMS data can be combined with data about health care pathways and used to evaluate the outcomes of care in terms of the improvement or deterioration in patients' quality of life including features such as their pain, mobility, ability to see and to hear, and ability to participate in regular daily activities.

Nine countries have 70% or more of the key national health and health care datasets

Ten countries reported 70% or more of these datasets are available at the national level: Canada, Denmark, Finland, Iceland, Japan, Korea, Norway, Singapore, Sweden, and Turkey (Table 2.1). Norway and Turkey were the only countries to report that national health datasets existed for all 14 types of personal health data. Other countries with strong data availability included Denmark (86%) and Finland, Iceland and Korea (79% each).

Virtually all countries responding to this study have some of the key datasets at the national level (Table 2.2). All countries reported national mortality data and virtually all reported national data for in-patient hospitalisations and mental health in-patient hospitalisations, a national population health survey and a national cancer registry. Virtually all also have a national population census or registry providing key population denominators and contextual data for health statistics and research. Sixteen of 22 countries reported national emergency health care data and prescription medicines data and 14 reported national data for primary health care. Just over half of countries reported national data for formal long-term care and ten reported national data about patient's care experiences. Eight countries reported a national registry for cardiovascular disease patients and five countries reported the same for diabetes patients. Patient-reported health outcomes data remain limited, with only four countries reporting these data exist at the national level. Further, only in the United Kingdom are these data incorporated into health care quality and system performance monitoring. Several countries signalled a broader range of available health data at the state or regional levels or within networks of health care organisations (see Box 2.1).

Box 2.1. Several countries have a broader range of health and health care quality linkage projects at the regional, state or health care organisation levels

Several countries signalled that a broader range of health data are collected at the regional or state level than is available at the national level. Sweden (12 datasets), Canada (10), Italy (10), Japan (9), New Zealand (9) and the United Kingdom (Scotland) (8) reported that the majority of datasets investigated in this study were available at the state or regional level and were being used in data linkage projects. Several datasets are also used in data linkage studies at the state or regional level in Spain (6), and Norway (5). The United States and the Netherlands also indicated that a number of datasets were available at the state or regional level and may be being used in data linkage studies. Denmark indicated that the cancer registry is available at a regional level.

Spain reported that a broader range of health and health care quality projects using record linkage are being conducted at the regional level than is possible at the national level. Spanish regions conducting health data linkage projects include Castilla-La Mancha, Baleares, Comunidad Valenciana, Cataluña, Extremadura, País Vasco, and Madrid. Reasons why included that there has been a strong decentralisation of health care to the regions; there are multiple public health research agencies that are region-based and region-financed; there are barriers to data availability at the national level; and there are monographic registries at the regional level that may lack legal or regulatory frameworks.

In Canada, health systems are organised and funded at the provincial level, as are health information systems and legislation regulating use of health information. Canada reported that provinces have developed datasets for laboratory tests, medical images and immunisations and have been linking these datasets with electronic medical and health records. Provincial centres for health data linkage in Canada include the Institute for Clinical Evaluative Sciences (ICES) in Ontario; the Manitoba Centre for Health Policy; the Newfoundland and Labrador Centre for Health Information; Pop Data BC for British Columbia and Alberta Health Services.

In Italy, regions conducting health data linkages include Lombardy, Piedmont, Friuli V.G, Emilia Romagna, Tuscany, Lazio, Veneto and Marche. Italy reported that regions have a broader range of data than is available nationally. The Ministry of Health, however, has a project called "Nuovo Sistema Informativo Sanitario (NSIS)" (i.e. New Healthcare Information System) that has the objective to collect individual health information for all care settings in all regions. By 2012, all regions had the ICT systems ready to collect the required data.

Many of the Swedish regions and counties are undertaking data linkage projects, for example, the region of Västra Götaland. Sweden reported that there are several diagnosis or procedure specific Health Care Quality Registries at the national level that are all run by county councils.

In Scotland, the Health Informatics Centre Dundee is actively engaged in health data linkage projects and in the United States, large states, such as California and New York likely engage in this type of work. Switzerland noted that a broader range of health care data is available within the cantons than is available nationally; however, no data linkage projects were reported.

Networks of health care organisations are also conducting data linkage projects with a broad range of health data in many countries. Networks of health care organisations, such as Kaiser Permanente in the United States, offer a broad range of health care services and can conduct research where patient data are linked across the different health care facilities they operate. The United States signalled that twelve of the 14 key datasets investigated in this study were available and used in data linkage studies within such networks. Similar levels of activity within networks were reported by New Zealand (ten datasets); Spain (nine datasets); and Israel, Japan and Singapore (seven datasets each). There was also activity reported in Canada (five datasets); Denmark (one dataset) and the Netherlands (one dataset).

Table 2.2. Dataset is available at the national level

	Hospital in-patient data	Mental hospital in-patient data	Emergency health care data	Primary care data	Prescription medicines data	Cancer registry data	Diabetes registry data	Cardiovascular disease registry data	Mortality data	Formal long-term care data	Patient reported health care outcomes data	Patient experiences survey data	Population health survey data	Population census or registry data
Canada	Yes	Yes	Yes	Yes	Yes	Yes	No	No	Yes	Yes	No	No	Yes	Yes
Czech Rep.	Yes	Yes [3]	No	No	No	Yes	No	Yes [6]	Yes	No	No	No	Yes	Yes
Denmark	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Finland	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	Yes	Yes
Iceland	Yes	Yes [3]	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	No	No	Yes	Yes
Ireland	Yes	Yes	No	No	Yes	Yes	No	No	Yes	No	No	Yes	Yes	Yes
Israel	Yes	Yes	Yes	No	No	Yes	No	No	Yes	Yes	No	Yes	Yes	Yes
Italy	Yes	Yes	Yes	Yes [4]	Yes	Ns	Ns	ns	Yes	Yes [5]	No	No	Yes	Yes
Japan	Yes	Yes	Yes	Yes	Yes	No	No	No	Yes	Yes	No	Yes	Yes	Yes
Korea (Rep. of)	Yes	Yes	Yes	Yes	Yes	Yes	No	No	Yes	Yes	No	Yes	Yes	Yes
Netherlands	Yes	Yes	ns	Yes	Yes	Yes	Ns	ns	Yes	Yes	ns	Yes	No	ns
New Zealand	Yes	Yes	Yes	No	Yes	Yes	No	No	Yes	No	No	No	Yes	Yes
Norway	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Singapore	Yes	Yes	Yes	Yes	No	Yes	No	Yes	Yes	Yes	No	No	Yes	Yes
Spain	Yes	Yes	No	No	No	No	No	No	Yes	No	No	No	Yes	Yes
Sweden	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
Switzerland	Yes	Yes	No	No	No	Yes	No	No	Yes	Yes	No	No	Yes	Yes
Turkey	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
United States	No [2]	No	Yes	Yes	Yes	Yes	No	No	Yes	No	Yes	Yes	Yes	Yes
UK England	Yes	Yes	Yes	No	No	Yes	Ns	ns	Yes	No	Yes	Yes	Yes	Yes
UK Scotland	Yes	Yes	Yes	Yes	Yes	Yes	No [1]	No	Yes	ns	No	No	No	Yes
UK Wales	Yes	Yes	Yes	Yes	Yes	Yes	Ns	ns	Yes	No	ns	ns	Yes	Yes
Total yes	21	21	16	14	16	19	5	8	22	12	4	10	20	21

1. National dataset exists and is in the custody of 14 Health Boards.

2. National hospitalisation data are in development.

3. Mental health in-patient data are part of the hospital in-patient dataset.

4. Prescription medicines given in outpatient care.

5. Includes residential, semi-residential and home care.

6. Includes surgery and similar procedures.

ns: Not stated.

Source: Authors own calculations based on the results of this study.

Progress in national dataset availability since 2011

There has been some progress in dataset availability among the twelve countries that participated in the OECD HCQI Information Infrastructure surveys in both 2011 and 2013. Canada is now reporting national datasets for prescription medicines and progress toward national data on primary health care;[2] and Switzerland is now reporting the availability of a national cancer registry.

Further, there are countries that are developing new sources of national data. The United States is developing national hospitalisation data and Canada and the Czech Republic[3] are developing national data on patient experiences. Switzerland's national cancer registry is now operational and a law is under development that will bring state participation in the national registry to 100%.

Further, several countries reported other key national health datasets. Japan reported national personal health data on intractable diseases is in development. Intractable diseases are defined as chronic, cause-unknown, untreatable diseases. Italy reported national data for hospice care and care for drug dependency. Spain reported national data for outpatient surgeries, hospital day care and ambulatory clinic visits. Birth notifications data were noted by the United Kingdom as a key dataset and Canada reported national data for public health surveillance of certain health threats.

The Czech Republic reported a loss of a key personal health dataset. Up to 2009, personal health data for prescription medicines were collected. However, the Office for Personal Data Protection determined that the State Office for Drug Control did not have the authority to collect the data. As a result, pharmacies now report prescription medicines data only by year of birth and sex. It is no longer possible to estimate prescription use at the level of individuals. A legislative reform has been initiated to enable a new registry with data provided by insurance companies.

Highest coverage of the target population in the key datasets of Denmark, Finland, Sweden and Iceland

The national personal health datasets reported by countries tend to have very high coverage of targeted populations (Table 2.3). Among the countries, virtually all of the datasets of Denmark, Finland, Sweden and Iceland were reported as having high coverage of the target populations, followed by those of Korea and Singapore (Tables 2.1 and 2.3).

Reasons for less than full coverage of national health datasets include federated countries where not all states or regions are yet participating (Canada, Switzerland); voluntary reporting systems [Japan, Canada, Netherlands, Norway and United Kingdom (Wales)]; datasets that cover only publicly provided or reimbursed services [Finland, Ireland, New Zealand, Netherlands, Singapore, Turkey and United Kingdom (England and Wales)]; datasets that are representative samples [United Kingdom (Scotland) for primary care and most countries for survey datasets]; datasets that include only certain patients, such as patients in a chronic disease management programme or disease-based health care quality register (Singapore and Sweden); and datasets that exclude certain care providers such as specialised hospitals (Israel); university hospitals (Turkey); veterans' health care and military hospitals (United States); and providers unable to submit patient-level data [United Kingdom (Wales)].

Table 2.3. Proportion of the population covered by the data

	Hospital in-patient data	Mental hospital in-patient data	Emergency health care data	Primary care data	Prescription medicines data	Cancer registry data	Diabetes registry data	Cardiovascular disease registry data	Mortality data	Formal long-term care data	Patient reported health care outcomes data	Population census or registry data
Canada	100%	100%	60%	< 5%	ns	100%	na	Na	100%	70%	na	100%
Czech Rep.	100%	100%	na	Na	na	100%	na	100%	100%	Na	na	100%
Denmark	100%	100%	100%	100%	100%	100%	100%	100%	100%	Na	na	Ns
Finland	100%	100%	na	100% [1]	100% [2]	100%	100% [7]	100%	100%	100%	na	100%
Iceland	100%	100%	ns	100%	100%	100%	na	100%	100%	100%	na	100%
Ireland	ns	100%	na	Na	77%	100%	na	Na	100%	Na	na	100%
Israel	75%	100%	100%	Na	na	100%	na	Na	100%	75%	na	100%
Italy	100% [3]	100% [3]	100% [3]	100% [3]	100% [3]	na	na	Na	100% [3]	100% [3]	na	100%
Japan	100%	100%	100%	94%	100%	na	na	Na	100%	90%	na	100%
Korea (Rep. of)	100%	100%	100%	100%	100%	100%	na	Na	100%	100%	na	100%
Netherlands	82% [4]	100%	na	3%	97%	97%	na	Na	100%	Ns	na	Ns
New Zealand	95%	100%	100%	Na	90%	100%	na	Na	100%	Na	na	100%
Norway	100%	100%	100%	100%	100%	100%	63%	100%	100%	Ns	ns	100%
Singapore	100%	100%	<100% [3]	<100%	na	100%	na	100%	100%	100%	na	<100%
Spain	100%	100%	na	Na	na	na	na	Na	100%	Na	na	100%
Sweden	100%	100%	100%	<100% [6]	100%	100%	100%	100%	100%	100%	na	100%
Switzerland	100%	100%	na	Na	na	68%	na	Na	100%	100%	na	100%
Turkey	66%	66%	66%	66%	66%	66%	66%	100%	100%	66%	66%	100%
United States	na	na	ns	Ns	ns	28%	na	Na	100%	Na	90% [5]	100%
UK England	ns	78%	ns	Na	na	100%	na	Na	100%	Na	100%	100%
UK Scotland	100%	100%	100%	6%	100%	100%	na	Na	100%	Na	na	ns
UK Wales	100%	100%	<100%	47%	100%	100%	na	Na	100%	Na	na	100%

1. Public services.

2. Reimbursed medications.

3. Public and/or nationally accredited institutions.

4. Short-stay hospitals in 2012. Up to 2004, coverage was 100%.

5. Approximate.

6. Within quality registers for certain diseases.

7. All patients with reimbursed medication for diabetes.

na: Not applicable. ns: Not stated.

Source: Authors own calculations based on the results of this study.

Automatic extraction of electronic data is prevalent in 13 countries

Thirteen countries reported that data are automatically extracted from electronic clinical and administrative data systems to populate 80% or more of their key national health care datasets (Table 2.1). Electronic data often must be complemented with other data sources, however, to complete datasets or to obtain a high population coverage. The majority of countries are populating some of their health care datasets by extracting data from electronic clinical records automatically (Table 2.4). This method can improve the timeliness and accuracy of the data and indicates that the medical records have been digitised.

In a significant share of countries there is manual entry of data to health care datasets from paper clinical records. In several countries there is a combination of both methods that is needed due to differences across providers in their reporting capabilities. A smaller number of countries reported that the source of health care data was automatic extraction or manual data entry from claim or billing records. Claim or billing records were most heavily depended upon for health care data in Korea, Norway, Japan, Singapore and Sweden. The United States and the United Kingdom (England) also rely on claim or billing records for hospitalisation data. Most countries rely on claim or billing records for prescription medicines data. Questionnaires are rarely used to collect national health care data.

Table 2.4. Number of countries reporting sources of variables within national datasets

	Hospital in-patient data	Mental hospital in-patient data	Emergency health care data	Primary care data	Prescription medicines data	Cancer registry data	Diabetes registry data	Cardiovascular disease registry data	Mortality data	Formal long-term care data
Data entry from paper clinical records	11	10	8	5	4	11	1	2	11	5
Data extracted automatically from electronic clinical records	16	17	12	12	7	12	4	6	7	7
Data entry from paper claim or billing records	5	3	4	2	7	1	1	1	2	3
Data extracted automatically from electronic claim or billing records	7	5	6	3	10	3	1	3	2	5
A survey questionnaire	2	1	1	1	1	0	0	0	4	2

Source: Authors own calculations based on the results of this study.

Twelve countries reported consistently coding health care data using a terminology standard

Half of the countries participating in this study reported that within all of their key health care datasets clinical terminology is coded by assigning standard codes using a classification system (Table 2.1). Standard codes ensure that data elements are comparable across datasets and can be analysed for statistical purposes.

Across countries, clinical terminology is coded by assigning standard codes using a classification system (Table 2.5). The majority of countries coding clinical terminology within hospital in-patient datasets, cancer registries and mortality datasets were doing so with the aid of health care coding professionals who have been trained to analyse clinical statements and assign standard codes. The use of such professionals was less frequently reported for other datasets. Instead, countries reported that health care professionals, such as nurses and doctors, assign the standard codes. The movement toward health care professionals entering and coding data accompanies the introduction of electronic medical and health record systems. It introduces data quality challenges and requires new approaches to ensure that data records are of high quality, such as health care provider training, data usability evaluations and auditing for data quality.

A previous OECD study explored in greater detail the data content standards that were being used for the coding of clinical elements within electronic health record systems (OECD, 2013a). It found considerable variety across countries in the terminology standards used with some countries adopting international terminology standards and others developing national standards. Further there were key data elements with no agreed international terminology standard. Progress toward internationally comparable indicators of health and health care from electronic clinical data will require greater harmonisation toward internationally-agreed terminology standards.

Table 2.5. Number of countries reporting coding clinical terminology

	Hospital in-patient data	Mental hospital in-patient data	Emergency health care data	Primary care data	Prescription medicines data	Cancer registry data	Diabetes registry data	Cardio-vascular disease registry data	Mortality data	Formal long-term care data
Clinical terminology is coded by assigning standard codes using a classification system	22	21	14	15	14	19	5	7	21	8

Source: Authors own calculations based on the results of this study.

Retention periods for personal health data

While many countries do not place a limit on how long national datasets containing personal health data can be held, some do so. In Korea, the national privacy law specifies the length of the retention period for each dataset and it does vary. Within the Health Insurance and Review Agency (HIRA), there are certain datasets with a very short retention period of one year and there are datasets with an indefinite retention period. For the major insurance claims datasets, the retention period is 70 years. After the retention period time limit is up, the dataset may be de-identified and archived outside of HIRA. After it is placed in the archive, it is legally permissible that the dataset could be retrieved and re-identified to enable an approved use.

In the Czech Republic, the custodian must preserve personal data only for the period of time that is necessary for the purpose of their processing. After expiry of this period, personal data may be preserved only for purposes of the state statistical service, and for scientific and archival purposes. When using personal data for these purposes, it is necessary to respect the right to protection of private and personal life of the data subject from unauthorised interference and to make personal data anonymous as soon as possible. At the ministry, data that contain unencrypted IDs may be lawfully held for five years. After five years, the data must be anonymised. The practice is to use a consistent method for the encryption of the ID numbers, so that the data can still be used in future approved data linkage projects.

In Israel, in the case of genetic information, there is a time limit for record retention. For medical records there is a minimum time medical records should be retained – but no upper time limit. In New Zealand, both the national *Privacy Act* and the Health Information Privacy Code (HIPC) provide that health information should not be kept for longer than is required for the purposes for which the information may be lawfully used. Health Information Regulations in New Zealand provide for a minimum retention period by practitioners of ten years since services were last provided, with no maximum retention period specified.

Concerns with the quality of the data

Countries were asked if they had any concerns with the quality of key national datasets that limit their usefulness.

The Czech Republic signalled that the national data collected by the Health Ministry (IHIS) is not linked to reimbursement decisions but is provided to IHIS from health care providers. There are no incentives for providers to be rigorous about the quality of the data submitted. The data verification processes at IHIS are routine logic checks similar to those applied by Eurostat. There is no capacity to validate the data by checking data records against original health care records. There is concern that particularly time-consuming aspects of the data requested from providers, such as the capturing of co-morbidities, may be of lower quality.

Iceland noted that frequently data are not coded in a timely manner and there is a lack of internal data quality audits within health care providers before data are submitted to the national authority. The Netherlands noted that missing data within datasets and the use of different coding systems for the same data elements are barriers to analysis. Norway notes that the lack of structured data and/or use of terminology standards for some data elements are barriers to quality and to analysis of the data.

Italy noted that difficulties harmonising data quality across its regions is a barrier to the usability of data at the national level. Spain expressed similar data quality challenges at the national level as well as gaps in the coverage of its national registries. There is also a need to advance data quality assurance standards in Spain.

In the United Kingdom, England signalled the lack of quality for certain data elements, such as the capturing of ethnicity within birth data.

Box 2.2. Electronic Health Record systems

The development and use of data from electronic health records (EHR) has the potential to support health care innovation and to improve the quality, safety and performance of health care systems. This is because such records can be brought together into an electronic health record system, which contains or virtually links together records from multiple care providers to create a longitudinal view of patients' health care pathways.

National EHR Plans

In 2012, most countries reported a national plan or policy to implement electronic health records (22 of 25 countries) and most had already begun to implement that plan by 2012 (20 countries) (OECD, 2013a; OECD, 2015). At that time, the implementation was relatively new in virtually all participating countries, having started within the previous four years. Of the 25 countries studied, 18 countries had included some form of secondary analysis of electronic health records within their national plan. The most commonly included secondary uses reported by 15 countries were public health monitoring and health system performance monitoring. Fourteen countries also indicated that they intended for physicians to be able to query the data to support treatment decisions. The least commonly-reported planned data use was for facilitating or contributing to clinical trials. This use was noted by ten countries. Many countries also reported that regular use of electronic health record data for secondary analyses were already underway. Public health monitoring (13 countries) and general research (11 countries) were the most commonly reported uses.

Key differences between countries envisaging data uses and those who are not

There are several significant differences between the 13 countries whose national plans or policies called for at least four different data uses (the engaged) and the twelve countries who were planning on fewer or no secondary data uses (the cautious). Engaged countries were somewhat more likely than cautious countries to report having created national governing bodies responsible for clinical terminology and interoperability standards, 62% compared with 50%. Terminology standards ensure that the data are captured in a consistent way with a structure that enables statistics and analysis. Interoperability standards ensure that records can be shared or exchanged.

Box 2.2. Electronic Health Record systems *(cont.)*

The majority of engaged countries (69%) are implementing an EHR system that will enable the sharing of records between and among physicians and hospitals and that will include information on current medications, lab tests and medical images. In contrast, none of the cautious countries are implementing an EHR system with all of these features. Further, virtually all of the engaged countries (92%) have developed a national minimum dataset that standardises the content of patient records that are intended to be shared among health care providers. In contrast, only one-half of the cautious countries have defined a minimum dataset. The majority of engaged countries, 62%, reported that all or most of the key data elements within their EHR (diagnosis, medications, lab tests, medical images and surgical procedures) follow clinical terminology standards. In contrast, only 17% of cautious countries have adopted clinical terminology standards to the same degree. When data are not coded to a terminology standard either at the point-of-care or after the fact, key clinical decision support algorithms such as reminders and alerts cannot be used by front-line clinicians, data cannot be shared effectively across care settings to support continuity of care, nor can it be analysed to monitor public health and health system performance and to conduct research (OECD, 2013a: CIHI, 2013).

Engaged countries (54%) are somewhat more likely than cautious countries (42%) to report that their EHR system is already being used to create datasets for statistics and research. Engaged countries are much more likely than cautious countries, however, to have put into place processes to evaluate the usability of EHR data for statistical purposes (69%, compared with 17%). As a result, it is perhaps not surprising that engaged countries (62%) are more likely than cautious countries (50%) to be concerned with the quality of the data being entered into electronic clinical records. Engaged countries (31%), compared with cautious countries (17%), are also more likely to have instituted processes for auditing the clinical content of electronic records for quality. Although, this is still relatively rare for both groups.

Six countries use all of their national health care datasets to regularly report about the quality and performance of health care

Among the countries participating in this study, Canada, Israel, Italy, New Zealand, Spain and the United Kingdom (Scotland) reported that all of their key national health care datasets are analysed to produce regularly reported indicators of health care quality or health system performance (Table 2.1).

Most countries reported developing indicators for health care quality or health system performance monitoring from hospital in-patient data (19 countries) and cancer registry data (18 countries); followed by mental hospital in-patient data (15 countries) and population health survey data (14 countries). About half of countries have gone further and are developing indicators across the continuum of health care.

Finland, Iceland, Singapore, Sweden, the United Kingdom (Scotland and Wales) have the highest proportion of key national health datasets sharing the same unique patient ID number

Most countries participating in this study have essential elements to develop health care pathway data. The majority have national datasets with records for patients or persons for at least hospital in-patient data, mental hospital in-patient data, cancer registry data, mortality data and population survey and census data (Table 2.6). Further, a majority of these datasets contain a number for each patient that uniquely identifies them and could be used for an approved data linkage, such as a social insurance number or a health insurance number. Others often have identifying variables, such as names, addresses and dates of birth that could be used to establish a dataset linkage.

Nonetheless, there is a problem among OECD countries in the consistency in which a common unique patient ID number is captured within key national health datasets. Iceland (100%) and Sweden (90%) clearly stand out in terms of the proportion of key national health datasets sharing the same unique ID, followed by the United Kingdom (Scotland) at 78% and Finland at 73% (Table 2.1).

Finland, Iceland, the United Kingdom (England) and Singapore are regularly linking most of their national health care datasets for statistics and research

Finland, Iceland, the United Kingdom (England) and Singapore reported that 90% or more of their key national health care datasets were being linked to other health and health care datasets to regularly monitor health care quality or system performance or to produce other approved statistics or research studies (Table 2.1).

Monitoring pathways of care and the outcomes that result is essential to understanding and improving health care quality and performance. Seldom is all of the data needed to understand a pathway available within a single dataset and data linkages are needed to bring together data across the continuum of care for the same patients. Dataset linkages also enable key characteristics of individuals, such as socio-economic status, health behaviours and environmental exposures to be linked to health care datasets to provide a new understanding about the impact of health risks and inequalities on individuals' health and health care.

Most countries reported developing indicators for health care quality or health system performance monitoring from hospital in-patient data (19 countries) and cancer registry data (18 countries); followed by mental hospital in-patient data (15 countries) and population health survey data (14 countries). About half of countries have gone further and are developing indicators across the continuum of health care.

Indeed, all of the countries with personal hospital in-patient data reported that the data are used for approved data linkage projects (Table 2.6). Similarly, the vast majority of countries reported data linkage projects involving their datasets across the spectrum of national health data, with the exception of survey data. Population health survey and patient experiences survey data are less likely to be reported as having identifying variables and only half of countries with these data reported having linked the data for approved projects.

Countries were asked if their data are linked on a regular-basis such that a project involving the linkage of the data is usually underway. Cancer registries were the most frequently reported as involved in a regular data linkage (17 countries), followed by hospital in-patient data (16) and mortality data (15).

Hospital in-patient data are a key input to most programmes of quality and performance monitoring. Among the 19 countries where hospital datasets contain a unique patient identifying number, 17 could link these data to mental hospital in-patient data, 15 to mortality data, 14 to population census or registry data, 13 to emergency health care data, 13 to cancer registry data, 11 to prescription medicines data and 10 to primary health care and long-term care data. This technical capacity to link is because the datasets share the same unique patient identifying number. Fewer countries link these data on a regular basis for health care quality monitoring. Very few countries include linkages of their mental hospital in-patient data, emergency health care data, primary care data, prescription medicines data, and long-term health care data in their programmes of health care quality and system performance monitoring.

Table 2.6. Just over half of countries with national datasets are regularly linking the data to monitor quality or health system performance

	Hospital in-patient data	Mental hospital in-patient data	Emergency health care data	Primary care data	Prescription medicines data	Cancer registry data	Diabetes registry data	Cardiovascular disease registry data	Mortality data	Formal long-term care data	Patient reported health care outcomes data	Patient experiences survey data	Population health survey data	Population census or registry data
Dataset contains records for patients (persons)[1]	20	18	14	10	14	19	7	8	21	11	7	8	19	19
Data set contains a number that uniquely identifies the patient (or the person) that could be used for an approved data linkage	19	18	13	10	12	16	6	8	17	10	4	5	12	14
Data set contains the same identifying number as is used for hospital in-patient data[2]	na	17	13	10	11	13	6	8	15	8	4	6	8	11
Dataset contains identifying variables (such as name, sex, birth date, address) that could be used to link these data to another set of data	16	15	11	9	11	15	5	6	17	10	4	5	11	13
Record linkage projects are conducted with these data	20	16	15	10	13	18	5	7	19	10	5	4	9	14
Record linkage projects are conducted with these data on a regular basis	16	13	10	4	10	17	2	5	15	7	4	2	7	9
Record linkage projects are conducted with these data on a regular basis for health care quality or system performance monitoring	13	7	6	2	3	12	1	4	11	2	1	0	2	2

1. Each row may be one treatment/visit/admission that can be grouped by person.

2. Where pseudonymisation algorithms are used for linkage, they would need to be consistent across datasets.

Source: Authors own calculations based on the results of this study.

Thirteen countries are regularly linking data from four national datasets: hospital and mental hospital in-patient data, cancer registry data and mortality data (Table 2.7). Ten of them are also regularly linking emergency care data, eight prescription medicines data, six long-term care data and three primary care data.

Table 2.7. Thirteen countries are linking data across the pathway of care

A	B	C	D	E
Regularly linking hospital in-patient, mental-hospital in-patient, cancer registry data and mortality data	Linking datasets in A + emergency care data	Linking datasets in A + prescription medicines data	Linking datasets in A + long-term care data	Linking datasets in A + primary care data
Canada	Canada	Canada	Canada	Korea
Czech Republic	Israel	Denmark	Finland	Singapore
Denmark	Korea	Finland	Israel	United Kingdom (Wales)
Finland	New Zealand	Korea	Korea	
Israel	Norway	New Zealand	Singapore	
Korea	Singapore	Sweden	United Kingdom (Wales)	
New Zealand	Sweden	United Kingdom (Scot. & Wales)		
Norway	United Kingdom (Eng., Scot. & Wales)			
Singapore				
Sweden				
UK (Eng., Scot. & Wales)				

Source: Authors own calculations based on the results of this study.

Countries provided examples of the purpose of the regular data linkages they are undertaking. Key reasons include to develop health care quality and system performance indicators including OECD quality indicators; to measure the co-ordination of care and health care pathways and outcomes; for estimates of compliance to national care quality guidelines; for indicators of health care utilisation and its cost; for measures of disease prevalence; and to measure health and health care use by socio-economic status. Linkage is also routinely necessary to assure the quality, completeness and validity of national datasets and to conduct medical and health services research projects.

These 13 countries also stand out for the number of national datasets regularly included in data linkage projects for health or health care monitoring, statistics or scientific research (Table 2.8). Iceland is also conducting data linkage projects with the majority of its key national datasets. Countries with few key national datasets regularly involved in a data linkage projects are Ireland, Italy and Switzerland. In Japan and Turkey there are no regular dataset linkages.

Table 2.8. Seven countries are linking seven or more key datasets on a regular basis for statistics or research

7+ key national datasets	Canada, Finland, Iceland, Israel, Korea, Norway, Singapore, Sweden, United Kingdom (England, Scotland and Wales)
5-6 national datasets	Denmark, New Zealand, Netherlands, United States
3-4 national datasets	Czech Republic, Spain
1-2 national datasets	Ireland, Italy, Switzerland
0 national datasets	Japan, Turkey

Source: Authors own calculations based on the results of this study.

Little change in data linkage activities since 2011

Twelve countries participated in the OECD HCQI studies in both 2011 and 2013. These are Canada, Denmark, Finland, Israel, Japan, Korea, Norway, Singapore, Sweden, Switzerland, United States and United Kingdom (England and Scotland). There has been some change in the number of countries reporting that national health datasets are involved in an approved data linkage project on a regular basis, such that a data linkage project is usually underway involving the dataset (Table 2.9).

By 2013, most countries were reporting data linkage projects on a regular basis with five key national datasets: hospital and mental hospital in-patients, deaths, registered cancers, and prescription medicines. There was some change with fewer countries reporting hospital in-patients and mortality data are routinely involved in linkage studies and a greater number indicating use of cancer registry, prescription medicines and mental-hospital in-patients data.

It remains the case that fewer countries are able to involve data about patients in primary care or long-term care settings in data linkage studies and the number of countries routinely linking primary care data to other data is lower in 2013. The number of countries indicating that data from a population census or registry, which provides socio-demographic information about the population, is routinely involved in data linkage projects has also fallen. About the same number of countries reported that data from population health surveys, which provide information on health risks and behaviours, is routinely involved in data linkage studies.

It is not possible to conclude change from this comparison, however, as the methodology used for the two studies has an important difference. The OECD study in 2013, conducted for this report, gathered detailed information about data governance practices and, as a result, survey participants in countries with multiple dataset custodians often consulted with them before responding. The burden of reporting was much lighter for the 2011 study.

Table 2.9. Number of countries[1] reporting a data linkage project is taking place on a regular basis involving national datasets in 2011 and 2013

Dataset	Hospital in-patients	Deaths	Cancers	Rx[2]	Mental hospital in-patients	Primary care	Long-term care	Population health survey	Census or population registry
A linkage study is usually underway in 2011	10	11	8	6	7	4	5	6	9
A linkage study is usually underway in 2013	10	11	12	7	10	2	5	5	7

1. 12 countries responded to the OECD survey in both years.

2. Prescription medicines data.

Source: OECD 2013a and authors own calculations based on the results of this study.

National projects advancing high-value data to promote health and improve health care

Countries that are actively monitoring health care quality and health system performance provide very interesting examples of the benefits of developing evidence-based management of health care systems.

- Finland monitors the content, quality and cost-effectiveness of a set of selected diseases and treatments (stroke, premature new-borns, hip fracture, breast cancer, schizophrenia, heart attack, hip and knee replacement surgery, and invasive heart surgery) by linking patient data for the Finnish population across the whole cycle of care from admission to hospital, to care by their community doctor, to the medications prescribed and deaths (OECD, 2013a). From both administrative data and data extracted from electronic health records they have new indicators for each hospital to evaluate treatment quality and cost including: mortality rates, emergency room visits and readmissions to hospital, infections and complications, and stays in nursing homes and home care visits. Hospital quality is improving as the results are publicly available.
- Korea uses population-wide health insurance claim data to identify underuse, overuse and misuse of therapies and to reduce variation in care practices through regularly reporting quality indicators including mortality and readmission after hospital procedures; inappropriate prescribing in primary care; and outcomes following discharge from mental health hospitals (OECD, 2013a). Korea links claims data for patients across the whole pathway of care and is able to report timely results.
- Japan has created a new medical insurance claims database to assist the Ministry of Health, Labour and Welfare in the preparation, implementation and evaluation of a plan to optimise medical care costs. Several cost and quality studies were undertaken and published as a special issue of the Journal of the National Institute of Public Health. These studies included a linkage of insurance claim data with data on the provision of guidance to patients during periodic health check-ups regarding metabolic disease (Okamoto et al., 2013). The study found a reduction in the onset of metabolic disease and in health care expenditures among patients who received guidance about reducing disease risk during health check-ups.
- Sweden is breaking new ground by using data to undertake both quality and efficiency assessments of clinical care guidelines (OECD, 2013a). These guidelines inform physicians and health care professionals about the most appropriate therapies for patients with different health profiles and problems. By following

patient's cycle of care they are able to evaluate the extent to which guidelines are being followed and whether the health outcomes of patients meet expectations or not. This evidence is then used to revise the guidelines, completing an on-going cycle of improvement in care quality and efficiency.

- Within the United Kingdom, England has a new initiative called care.data that aims to create data about episodes of care including both health care and social care and involving pathways between primary and secondary care and information about diagnosis, laboratory tests and prescription medications (NHS, 2013). The six aims of the care.data initiative include supporting patient's choice, advancing customer services, promoting greater transparency, improving outcomes, increasing accountability and driving economic growth by making England a centre for world-class health services research. Data will be linked for consenting patients within the whole population of England, with data extracts taking place monthly to ensure timely monitoring.
- England has also concentrated the collection and linkage of large national personal health databases. This includes the new Health and Social Care Information Centre as the single national repository of health data and the Clinical Practice Research Datalink which provides access to data from electronic records for primary care doctors and facilitates linkages with other data, such as clinical trial cohort data and data in the custody of the Health and Social Care Information Centre (OECD, 2013a).
- The United States Food and Drug Administration has implemented a sentinel project to transform how it monitors the safety of the medicines, medical devices and biologics that it regulates by tapping directly into electronic health records, administrative data and insurance claim records. Building toward a nationwide rapid-response electronic safety surveillance system, the sentinel pilot study involves 17 data partners across the United States, and encompasses the data of nearly 100 million patients (FDA, 2013).

Countries provided 27 examples of current or recently completed national projects involving the linkage of datasets or the extraction of data from electronic clinical records to follow the health care pathway and assess health care outcomes; and to understand how outcomes vary by socio-demographic characteristics, health-risk behaviours and health conditions. These projects further illustrate the potential for routinely collected data to improve our understanding of what works, for which patients, when and at what price.

1. Canada: Electronic Medical Record and Electronic Health Record Proof of Concept Project

Purpose	To demonstrate the use of electronic medical record (EMR) and electronic health record (EHR) data for research and other secondary uses by conducting three studies: sex differences in risk factors for adverse outcomes in diabetes; effects of obesity on health care services utilisation and chronic disease; and psychiatric medication adherence and its relationship with hospital re-admissions.
Organisations involved	Newfoundland and Labrador Centre for Health Information and the Canadian Centre for Health Information
Data involved	Electronic Medical Records (EMR) and Pharmacy Network data were linked to three administrative health databases (i.e., hospital, physician claims and mortality data).
Description	The project is investigating the feasibility of using data from EMR and EHR systems in the Canadian province of Newfoundland and Labrador for health research and health systems uses. The project is documenting the processes and factors associated with the use of EMR and EHR data including challenges and lessons learned. It is intended that project findings will serve as a model and facilitate future research and health system use involving EMR and EHR data, as well as contribute to the ongoing development and evolution of these systems so that their full benefits as data sources can be realised. The project has identified key factors that are important to consider when utilising EMR and EHR data for research. These have been categorised as: governance, approvals, data processing, and adoption.
To learn more	Newfoundland and Labrador Centre for Health Information (2013), E-Health Backbone, http://www.nlchi.nl.ca/newsletters/2013/february/

2. Canada: Canadian Primary Care Sentinel Surveillance Network (CPCSSN)

Purpose	Canada's first multi-disease electronic record surveillance system.
Organisations involved	The initiative is funded by the Public Health Agency of Canada under a contribution agreement with the College of Family Physicians of Canada on behalf of ten practice based research networks (PBRNs) associated with departments of Family Medicine across Canada. CPCSSN also works together with the Canadian Institute for Health Information.
Data involved	It collects and maintains national epidemiological surveillance data using Electronic Medical Records (EMRs) to improve outcomes in primary health care.
Description	The information gathered will help physicians to better understand chronic diseases and to improve the care provided to Canadians with chronic diseases and will support better management of health care systems. The Canadian Institute for Health Information is working with Canada Health Infoway and other partners to promote voluntary content standards for EMR systems in Canada that will enable the CPCSSN initiative to grow and will support health system uses of data through the CIHI Primary Health Care Voluntary Reporting System (Webster et al., 2011).
To learn more	The CPCSSN website is http://cpcssn.ca/. Project results are featured monthly in the Sentinel Eye section of the *Canadian Family Physician Journal* published by the College of Family Physicians of Canada, http://www.cfpc.ca/CanadianFamilyPhysician/. Recent articles have reported on how CPCSSN data are improving pharmacovigilance, which is the reporting of adverse drug reactions, and the need to develop national content standards to support analysis of data from EMR records (Keshavjee et al. 2014; Williamson et al. 2014).

3. Canada: Innovations in Data, Evidence and Applications for Persons with Neurological Conditions (ideas PNC)

Purpose	To understand the strengths, preferences and needs of persons with neurological conditions across the continuum of care.
Organisations involved	Project funded by the Public Health Agency of Canada and involving data linkages conducted by the Canadian Institute for Health Information and the Ontario Institute for Clinical and Evaluative Studies
Data involved	Data from inter RAI assessments within long-term care, home care and mental health care datasets were linked to data on in-patient hospitalisations, emergency hospital care, pharmaceutical data and primary health care data.
Description	The project aims to identify factors influencing the trajectory of change, quality of life, and health service utilization patterns across the continuum of care for patients with different neurological conditions including brain injuries and Alzheimer's disease and dementia.
To learn more	A description of the project and initial findings is available through the following link: http://interraicanada.uwaterloo.ca/iPNC/knowledge-bank/presentations/innovations-in-data-evidence-and-applications-for-persons-with-neurological-conditions/

4. Czech Republic: Project to Support Provision of Social Services

Purpose	Project to support the availability of social services for individuals with long-term care needs. The socio-economic and health needs of individuals in receipt of care allowances were studied in order to determine how best to organise long-term care services to meet their needs (Daňková et al., 2011).
Organisations involved	Project of the Ministry of Employment and Social Affairs.
Data involved	Data about individuals in receipt of care allowances was linked to survey and health insurance administrative records.
Description	The project evaluated the socio-economic profile, health services use and limitations in activities of daily living of the population in receipt of care allowances in order to determine how long-term care could be better organised to meet their needs.
To learn more	Study results are available in the Czech language: http://podporaprocesu.cz/wp-content/uploads/2013/01/Analyza_prijemcu.pdf

5. Denmark: Sundhedskvalitet (Health Quality)

Purpose	Produce a platform where citizens who need care in a hospital can access a set of performance indicators for individual hospitals.
Organisations involved	Project of the Statens Serum Institute.
Data involved	Hospitalisation data disaggregated by hospital unit with linkage to calculate quality indicators.
Description	Indicators are provided by hospital regarding waiting times for procedures and indicators of the quality of hospital services by disease or procedure such as lengths of stay, re-admission rates and re-operation rates.
To learn more	Indicators are available in the Danish language: http://www.esundhed.dk/sundhedskvalitet/Sider/sundhedskvalitet.aspx

6. Iceland: Patient safety – The incidence of adverse events and medical errors in Icelandic hospitals

Purpose	To investigate whether the incidence of adverse events and medical errors in hospitals in Iceland are similar to the findings of similar studies in neighbouring countries.
Organisations involved	Project of the Directorate of Health.
Data involved	Data gathering was done through retrospective reviews of 1000 case records that were randomly drawn from hospital records in 2009.
Description	The study is conducted in two hospitals in Iceland that represent around 85% of hospital activity in Iceland. The first part of the study has been completed and the results are similar to those of neighbouring countries.

7. Iceland: Non-adherence to prescribed medication in general practice

Purpose	To determine the prevalence of non-adherence to prescribed medication in general practice and to test whether it has been influenced by the moderate increase in patient co-payment that was implemented in 2010 (Linnet et al. 2013). Differences between co-payment groups were examined.
Organisations involved	Centre of Development, Primary Health Care of the Capital Area, Reykjavik, Iceland
Data involved	A population-based data linkage study. Prescriptions issued electronically by 140 physicians at 16 primary health care centres in the Reykjavik capital area were matched with those dispensed in pharmacies, the difference constituting primary non-adherence (population: 200 000; patients: 21 571; prescriptions: 22 991).
Description	The study examined prescriptions issued before and after the introduction of a moderate co-payment. Eight drug classes were selected to reflect symptom relief and degree of co-payment. Two-tailed chi-square test and odds ratios for non-adherence by patient co-payment groups were calculated. Primary non-adherence in Icelandic general practice was within the range of prior studies undertaken in other countries and was not adversely affected by the moderate increase in patient co-payment.
To learn more	http://www.ncbi.nlm.nih.gov/pubmed/22964077

8. Ireland: CaPPE – Cancer Pharmacoepidemiology & Pharmacoeconomics

Purpose	Investigation – at the population-level – of the effects of medications on cancer and assessment of the economic impact of medications and drugs in cancer. An example of results from this project is a study investigating associations between metformin exposure and colorectal cancer–specific survival using population-level data (Spillane et al., 2013).
Organisations involved	CaPPE is a collaboration between the National Cancer Registry, the Department of Pharmacology & Therapeutics, Trinity College Dublin and the National Centre for Pharmacoeconomics.
Data involved	This project uses a linked dataset consisting of cancer registrations and prescription data from Primary Care Reimbursement Scheme.
Description	The focus of the collaboration is research in the areas of cancer pharmacoepidemiology and pharmacoeconomics. Pharmacoepidemiology research involves the investigation – at the population-level – of the effects of medications on cancer. Pharmacoeconomics assesses the economic impact of medications and drugs in cancer.
To learn more	http://www.ncri.ie/research/scientific-papers/cohort-study-metformin-exposure-and-survival-patients-stage-i-iii

9. Israel: Psychiatric hospitalisations and deaths

Purpose	To investigate the overall mortality and selected natural and external causes of death by age, sex and mental health-related variables among persons who were ever admitted to psychiatric inpatient services (Haklai et al., 2011).
Organisations involved	Ministry of Health, Israel
Data involved	The national database on causes of death was linked to the mental health hospitalisation registry.
Description	This cohort study compared the mortality risk among persons aged 18 and over who were ever hospitalised in psychiatric facilities until 2006 with never hospitalised subjects. Mortality rates were computed by age, sex and psychiatric diagnosis, while proportions of deaths were computed by time from discharge. Rates were also analysed by time periods of date of death to check for possible associations with mental health policy decisions. Age-adjusted and age-specific mortality rates and rate ratios (RR) were computed for persons with a mental health hospitalisation compared with those who were never hospitalised. The study found that the age-adjusted mortality rate of hospitalised psychiatric persons was double than that of the non-hospitalised. The rate was higher in both sexes and for persons of all age groups, particularly for the young. The highest rate ratios were found for external causes of death, in particular suicide.
To learn more	http://www.questia.com/library/1P3-2577025031/the-mortality-risk-among-persons-with-psychiatric

10. Israel: Infant mortality

Purpose	To analyse infant mortality (Haklai et al., 2010).
Organisations involved	Ministry of Health, Israel
Data involved	The death database was linked to the live births database.
Description	Analysis of infant mortality rates, by cause, socio-demographic characteristics of the mother, and by particular types of high-risk infants, including pre-term births and infants with a low birth weight.
To learn more	This link presents the study findings in multiple languages including English: http://www.health.gov.il/PublicationsFiles/HealthIsrael2010.pdf

11. Italy: New Healthcare Information System

Purpose	To build an integrated system of homogeneous, individual health care information records, where patient information and the care delivery structure are the central information items. The goal is to make information available on the operating facilities at all health care levels, the services delivered, the resources used, and the related costs.
Organisations involved	Directorate General for Health Information and Statistical System - Department of Planning and Organization of the National Health Service, Ministry of Health
Data involved	Project involves the linkage of individual data across the spectrum of health care encounters in order to understand health care pathways, costs and outcomes.
Description	The project, on-going since 2006, is to develop a national data warehouse in which information necessary to support a strategy to balance costs and quality in the National Health Service is included. The project is to contribute significantly to public health authorities' governance capabilities by ensuring that the required analytical data on individual's health care pathways is available, including methodologies and rules to measure efficiency, appropriateness and costs overall and for different levels of government. The project is authorised by a Ministry of Health Regulation that was developed with the advice of the Italian Data Protection Authority.
To learn more	This link presents the study in Italian: http://www.nsis.salute.gov.it/

12. Italy: National Programme Outcome Evaluation (PNE)

Purpose	The programme brings together data from sub-national levels to evaluate the effectiveness and efficacy of health care services as well as providing a comparative evaluation of health care services.
Organisations involved	National Agency for Regional Health Services (AGENAS)
Data involved	Data on hospital admissions and discharges is linked to itself to determine hospital pathways and to data on deaths ascertained from the tax registry.
Description	The main objectives of the programme which was initiated in 2013 are to evaluate the efficacy of health interventions where there is a lack of clinical trial results; to evaluate the effectiveness of new treatments and technologies in real-world populations; to compare outcomes of health care services and to enable results to support accreditation, provider payments and public information; and to evaluate equity in health care availability and outcomes by comparing results for socio-demographic and regional population groups and to contribute results for accreditation and auditing.
To learn more	This link presents the study in Italian: http://www.salute.gov.it/portale/temi/p2_6.jsp?lingua=italiano&id=2905&area=programmazioneSanitariaLea&menu=vuoto

13. Japan: Health Data Project

Purpose	The project aims to generate evidence to support extending healthy life expectancy and to reduce future medical expenditures (Kumakawa et al., 2013).
Organisations involved	Ministry of Health, Labour and Welfare in collaboration with health insurers.
Data involved	Data on health insurance claims are linked at the individual level to establish health care pathways and to data on health check-ups required by insurers.
Description	The government is working with health insurers to link and analyse data from health insurance claims and health check-ups to implement efficient and effective health services for their subscribers. The project includes developing evidence to support population approaches to health care services to maintain and enrich health; and approaches to providing services to the population at high-risk of developing non-communicable diseases (NCDs), as well as health care approaches to improving outcomes for those diagnosed with NCDs.
To learn more	This link presents findings from this project in Japanese with summary information in English: http://www.niph.go.jp/journal/data/62-1/e62-1.html

14. Korea: HIRA Quality Assessment

Purpose	To improve health care quality by assessing the adequacy of medical services and inducing health care providers to steadily improve services found to be inadequate based on the assessment outcome.
Organisations involved	Health Insurance Review and Assessment Service (HIRA)
Data involved	Data across the spectrum of health care services are drawn from health insurance claim records and are linked to develop health care pathways and assess their outcomes. For some assessments, such as those for the treatment of AMI patients and the CABG procedure, a population selected from the medical claims data warehouse is sent to medical care institutions and information on clinical care is provided back via a web-based data collection system. For the outcome of death following health care services, hospital in-patient data are linked to computerised resident registration data from the Ministry of Security and Public Affairs. Prescription data are also linked to assess quality. For example, atypical anti-psychotics and readmissions within 30 days of hospital discharge for schizophrenia are calculated by linking mental hospital in-patient data to data on prescription medicines and hospital in-patient data.
Description	The results of the assessments are made public on HIRA's website for individual healthcare institutions so that the public can use the assessment results when choosing healthcare services. Healthcare institutions are notified of the assessment results and HIRA supports their quality improvement process through consultations. Overall assessment results are reported to the government for use in policy decisions. Assessment results for items included in a Value Incentive Program, including acute myocardial infarction and Caesarean delivery are sent to insurers in order to increase or reduce the reimbursement rate for the services provided. The number of items assessed is growing over time as is the proportion of total health expenditures that are assessed. In 2012, almost 40% of total health expenditures were included in the quality assessment.
To learn more	A comprehensive report of the methods and findings is available in English: http://www.hira.or.kr/eng/news/01/__icsFiles/afieldfile/2013/09/11/Comprehensive_Quality_Report_2012_eng.pdf

15. New Zealand: Health Quality and Safety Indicators

Purpose	The overarching goal for the indicators project is to develop a set of national health quality and patient safety indicators that support improvement of health services in New Zealand by comparing results within the country and comparing New Zealand to other countries.
Organisations involved	Health Quality and Safety Commission
Data involved	Hospitalisation data as well as mortality data and survey data sources are used. There is linkage within hospitalisation data to examine re-admissions.
Description	A set of health quality and safety indicators are published, as is an atlas showing variation in the quality of health care received by people in different geographical regions. There are quality and safety markers to track and incentivise progress in four critical areas of safety and quality: reducing harm from falls, hospital-acquired infections, and surgery and medication safety. Quality accounts are currently being adopted where health care providers account for the quality of their services in a similar way to financial accounts that show how an organisation used its money.
To learn more	http://www.hqsc.govt.nz/our-programmes/health-quality-evaluation/

16. Norway: Social Inequalities in Health

Purpose	Project to evaluate social inequalities in health.
Organisations involved	Norwegian Institute of Public Health
Data involved	Census data and data from the tax register on household income have been linked to the death register. A multi-generation database of the Norwegian population has been developed to enable exploration of cohort effects.
Description	The project is working to describe trends in social inequality in Norway from 1960 to the present. To date, the project includes analysis of mortality in children and adults and life expectancy. This includes a project to examine inequality in twenty different causes of death among adults and children (Næss Ø et al. 2007). In this study, almost 800,000 Norwegians in the age group 0–20 years in 1960 and still alive in 1990 were followed for deaths from 1990 to 2001. Results identified a set of causes of death in men and in women that showed a gradient by socio-economic status. Current research includes a study of the impact of factors in early life and across generations and their interaction with later lifestyles to explain social inequalities for a number of important common diseases in adulthood. The impact of the residential history (mobility) and individual socioeconomic position throughout the life cycle on geographical variation in mortality is also being studied.
To learn more	Information on the research programme on socio-economic inequalities and health is available in English http://www.fhi.no/eway/default.aspx?pid=240&trg=MainContent_6894&Main_6664=6894:0:25,7630:1:0:0:::0:0&MainContent_6894=6706:0:25,7754:1:0:0:::0:0&List_6673=6674:0:25,7649:1:0:0:::0:0:

17. Singapore: Retirement and Health Study

Purpose	To provide government agencies with a longitudinal database to better understand changes to Singapore's health and retirement landscape.
Organisations involved	Jointly conducted by the Central Provident Fund Board, the Housing and Development Board, the Ministry of Finance, the Ministry of Health and the Ministry of Manpower.
Data involved	A longitudinal survey of 25,000 persons aged 45 to 85 with the first wave collected between July 2014 and March 2015. The same individuals will be surveyed every two years over a ten-year period. Data are collected during the interview and, with consent, is retrieved from government databases including healthcare expenditures.
Description	The study aims to yield data and insights that enrich the policy making process by allowing interactions between employment status, wealth, health and retirement adequacy to be analysed.
To learn more	http://mycpf.cpf.gov.sg/Members/Gen-Info/RHS_2014.htm

18. Spain: Quality Indicators

Purpose	To study the quality of inpatient hospital services
Organisations involved	Ministry of Health, Social Services and Equality
Data involved	The Minimum Basic Data Set (MBDS) from the electronic record system provides over 4 million cases each year that could support national quality indicators.
Description	A set of national quality indicators is being constructed from MBDS data. A data repository and protocols for data analysis have been developed. Researchers have analysed the MBDS data to generate quality indicators for Spanish regions. For example, the MBDS of the Autonomous Region of Madrid was analysed to generate a suite of patient safety indicators following those of the US Agency for Health Care Research and Quality (Merchante et al., 2010).

19. Sweden: Open Comparisons of Quality and Efficiency in Swedish Health Care

Purpose	To compare health care quality across counties and hospitals
Organisations involved	National Board of Health and Welfare and the Swedish Association of Local Authorities and Regions
Data involved	Data are from many sources with indicators developed for quality of care outcomes resulting from data linkages.
Description	An annual report, published in November, provides a series of indicator-based comparisons of healthcare quality and efficiency among the various regions and counties in Sweden (Socialsyrelsen, 2013a). One purpose of the report is to make the publicly financed healthcare system more transparent. Another purpose is to promote healthcare management and control. The most recent report presents results for about 170 different indicators reflecting various dimensions of quality and efficiency concerning the healthcare system in general, as well as for different types of diseases.
To learn more	This link is to the indicators publication in English: http://www.socialstyrelsen.se/publikationer2013/2013-5-7

20. Sweden: National Performance Assessments

Purpose	To compare compliance to national health care guidelines across counties and hospitals.
Organisations involved	National Board of Health and Welfare
Data involved	Patient registries for each focussed condition for assessment are developed and then linked to relevant data such as mortality data and data on prescription medicines.
Description	There is one project for each national guideline for clinical care: Stroke care, Heart care, Diabetes care, Cancer care, etc. For the assessment of stroke care, for example, 44% of municipalities were found to collaborate with county councils or to have procedures in place for collaboration regarding rehabilitation of stroke patients (Socialstyrelsen, 2013b). Areas with collaboration had more favourable outcomes for patients. However, variations between municipalities are considerable and stroke patients' mortality rates and ability to manage activities in daily life have only improved marginally over the past ten years. Based on this assessment, the National Board of Health and Welfare has identified a number of areas for improvement of stroke care including reducing waiting times, increasing care provided in designated stroke units and improving rehabilitation care during and after hospitalisation.
To learn more	Example on stroke care is available in English: http://www.socialstyrelsen.se/publikationer2013/2013-3-4

21. Switzerland: Swiss National Cohort

Purpose	A multi-purpose, census based cohort and research platform permitting a better understanding of socio-economic and demographic characteristics of mortality and life expectancy (Spoerri et al., 2010).
Organisations involved	Federal Statistical Office with a consortium of university researchers.
Data involved	Data from the population census is linked to mortality data and this research platform may then be linked to other data sources for approved projects, such as cancer registries, live births, hospital stay episodes, and other clinical or population-based cohort studies.
Description	The cohort supports research to monitor and explain the evolution of socio-economic inequalities in mortality and survival outcomes in Switzerland. It is used to monitor and explain mortality differentials, while taking into acfcount individual socio-demographic, household and area-based characteristics.
To learn more	This link is to comprehensive information about the cohort and research results in English: http://www.swissnationalcohort.ch/index.php?id=2978

22. United Kingdom England: General Practice Extraction Service

Purpose	To develop national patient-level data for primary health care services in England for quality monitoring, provider payments and statistics and research.
Organisations involved	Health and Social Care Information Centre
Data involved	The service ensures that information from general practice systems in England can be accessed and used efficiently in a standardised way. Data are extracted from general practice IT clinical record systems.
Description	Data are extracted to support statistics and research including monitoring physician performance against the requirements of quality and outcomes frameworks and provision of payments to GPs and clinical commissioning groups. Data are available to the NHS and other approved organisations.
To learn more	http://www.hscic.gov.uk/gpes

23. United Kingdom England: Secondary Uses Service

Purpose	A data warehouse containing patient-level information used by NHS providers and commissioners for monitoring and research purposes (other than primary clinical care).
Organisations involved	Health and Social Care Information Centre
Data involved	Acute care hospitalisation data including in-patient data, out-patient data and emergency care data. Acute care data are also linked to national death data for indicators of survival.
Description	The data warehouse is analysed for healthcare planning, commissioning services, pay-for-results programs, improving public health and developing national policy. A range of services are provided to support data analysis, reporting and presentation including key performance indicators and data quality dashboards.
To learn more	http://www.hscic.gov.uk/sus

24. United Kingdom Scotland: Extending the understanding the impact of diabetes in Scotland

Purpose	A project to support continuing the Scottish diabetes register data linkage
Organisations involved	The Scottish Diabetes Research Network (SDRN) Epidemiology Group which receives funding from the Chief Scientist Office, Scottish Government.
Data involved	The Scottish Care Information – Diabetes Collaboration (SCI-DC) dataset forms an electronic, population-based register of over 99% of people with a diagnosis of diabetes in Scotland. It is linked to other routine databases, such as hospital admissions and deaths.
Description	The linked data supports research projects and clinical care auditing. As an example, the diabetes collaboration dataset was linked to hospital admissions and deaths to study the risk of cardiovascular disease (CVD) and total mortality among patients with type 1 diabetes (Livingstone et al. 2012). It found that the relative risks for CVD and total mortality associated with type 1 diabetes have declined relative to earlier studies, but continue to be associated with higher CVD and death rates than in the non-diabetic population.
To learn more	Information about the Scottish Diabetes Research Network is available here: http://www.sdrn.org.uk/node

25. United Kingdom Scotland: Growing Up in Scotland Survey Data Linkage

Purpose	Growing Up in Scotland is a longitudinal research study tracking the lives of thousands of children and their families from the early years, through childhood and beyond to study a range of social, educational and developmental risk factors and outcomes including physical and mental health and wellbeing.
Organisations involved	The survey is managed by ScotCen Social Research and is supported by the Scottish Government and the Medical Research Council.
Data involved	A large scale multi-cohort longitudinal survey of children living in Scotland. The survey is linked to health and education records with consent.
Description	Survey data are linked to health records and made available for research into public health issues such as infant feeding and child growth, and exposure to second hand smoke and respiratory illnesses.
To learn more	Information on the survey is available here: http://growingupinscotland.org.uk/

26. United Kingdom Wales: Flying Start Data Linkage Demonstration Project

Purpose	Flying Start is an early years intervention programme of the Welsh Government to improve the life chances of children in disadvantaged communities. This demonstration project was undertaken to establish whether the impact of the Flying Start programme could be evaluated from analysis of linked health and education administrative datasets.
Organisations involved	Welsh Government
Data involved	Flying Start areas are school catchments in deprived areas across Wales. Households in Wales were identified as belonging to the Flying Start areas, to the next most deprived areas and to other areas in the rest of Wales. Residences with children under age four in each year from 2004 to 2012 were selected for linkage to annual administrative data including hospitalisations, primary care, immunisation, educational outcomes and others.
Description	For each of the three groups of areas, variation in demographic characteristics and health and education indicators are compared and can be followed up over time. A set of indicators were estimated to support programme evaluation including those related to breastfeeding, immunisation, early hospitalisations, respiratory conditions, infectious diseases, and injuries.
To learn more	A complete description of the project, first results and limitations is available here: http://dera.ioe.ac.uk/19824/1/140131-data-linking-demonstration-project-flying-start-en.pdf

27. United States: Linked health survey and administrative data research platform

Purpose	To create a platform for research studies through the linkage of population health survey data and administrative data.
Organisations involved	National Centre for Health Statistics (NCHS)
Data involved	NCHS population health surveys are linked to death records and to medical insurance claims data from the Medicare and Medicaid programmes.
Description	Linked data files enable researchers to examine the factors that influence disability, chronic disease, health care utilization, morbidity, and mortality. The NCHS standardises and prepares population health survey records for linkage to health care and death records for approved projects. Examples of research results from analysis of the linked data include a study of the relationship between suicide and family status (Denney, 2010) and a study identifying chronic conditions in insurance claims data (Gorina et al., 2011).
To learn more	Information on the NCHS data linkage service is available here: http://www.cdc.gov/nchs/data_access/data_linkage_activities.htm

Key international projects to improve health care outcomes, safety and performance

Investments in the development of internationally comparable population-level health data are leading to new ways to benchmark and compare how health systems are performing to help countries to improve patient safety, health outcomes and system performance. For ten years, the *OECD Health Care Quality Indicators Programme* has been developing and reporting indicators of quality and performance across the domains of primary care, patient safety, hospitalisation outcomes and cancer care. This collaborative initiative has resulted in progress in the methodologies for comparable indicators, as well as progress in the development of the underlying data that enable the indicators. As of 2013, however, only one-half of OECD countries were able to report quality indicators requiring dataset linkages, such as mortality within 30 days after hospital admission for AMI or for Ischemic stroke (OECD, 2013b). Only seven countries were able to report on excess mortality from schizophrenia or from bipolar disorder.

Within Europe there are collaborative efforts funded by the European Union to advance health system performance and quality through analysis of large-scale databases. A few key examples from the EU seventh framework research programme are EU-ADR, EuroHOPE, and ECHO.

- The *EU Advanced Drug Reportin*g (EU-ADR) initiative defined a proactive strategy for post market drug assessment based on automating analysis of data stored in large electronic health record databases in four European countries (Denmark, Italy, the Netherlands, and the United Kingdom) and covering 30 million patients (Coloma et al., 2012). EHR data are analysed to identify a ranked list of identified signals of potential adverse events and their significance in terms of health risks. Adverse events monitored include acute myocardial infarction, acute renal failure, anaphylactic shock, and gastrointestinal bleeding. Results indicate that active surveillance with health care database networks for signal detection are feasible but an expansion of the coverage of the data network to a larger pool of patients, that is to more participating countries, would be necessary to monitor the effects of infrequently used drugs.
- *EuroHOPE, the European Health Care Outcomes, Performance and Efficiency Project* is evaluating the performance of European health care systems within seven countries in terms of outcomes, quality, use of resources and costs (Häkkinen et al., 2013). Participating countries include Finland, Italy, Netherlands, Norway, Sweden and the United Kingdom (Scotland). Health care data for hospitalisations, pharmaceuticals, registered cancers and deaths are linked to follow patient pathways of care. The patient groups studied are those with acute myocardial infarction, stroke, hip fracture, breast cancer and low-birth weight. EuroHOPE is developing indicators that it will recommend to the European Union for routine reporting; developing methods for international comparative health services research based on the linkage of person-level data; and informing about the policy-relevant drivers of health care quality including treatment practices, use of medicines and new medical technologies, waiting times, organisation of care and costs.
- *ECHO, the European Collaboration for Healthcare Optimisation* project, has pooled hospital administrative and contextual data from seven countries (Austria, Denmark, England, Portugal, Slovenia, Spain and Sweden) to learn

more about variation in access to care and outcomes of care and the relationship between this variation and the socio-economic status of the areas in which patients live (ECHO, 2013). ECHO intends to explore whether place of residence and access to particular health care providers influences access to safe and effective care by examining within-country and between-country variations. ECHO is the first international health system performance comparison where personal health data have been pooled, resulting in a dataset of 200 million hospital discharges.

In terms of single disease areas, there is no doubt that the long tradition of the development of databases to register cases of cancer has privileged cancer research with the evidence necessary to monitor and advance quality of care. There is also a long history in most OECD countries of linking cancer registrations and death databases to estimate cancer survival rates. *The International Cancer Benchmarking Project* is advancing this research further. In this project, cancer registries with detail about cancer stage at diagnosis have been analysed in six countries (Australia, Canada, Denmark, Norway, Sweden and the United Kingdom) to compare differences in survival and to discover why differences occur. Thus far, the researchers have found that patients in Sweden are the most likely to survive at least one year after diagnosis of breast, bowel and lung cancers; while those in the United Kingdom are the least likely (Cancer Research UK, 2013). The role of treatment in survival differences by cancer stage is the next stage of inquiry for the project.

Projects building platforms for internationally comparative statistics and research from data linkages and extraction of data from electronic clinical records

Title	Description	Participating countries	To learn more
The Farr Institute at the Centre for Improvement in Population Health through E-records Research (CIPHER)	To promote research using data from electronic health records and the linkage and analysis of large health-related datasets including social, economic and spatial data; and to build multi-disciplinary capacity in e-health information research. The project is developing collaborations to link previously isolated silos of expertise (observational, interventional, biomedical and social science); improving knowledge exchange between academic, practitioner and policy leads; and liberating information trapped in data islands. The project aims to enable routine health care data to be maximised by enabling research on the full UK population and robust methods to link such data to UK cohorts, surveys and non-health administrative data including embedding cohorts, trials and survey data within this total population structure. The project is to provide the data, methods and skills to enhance observational and interventional research capacity and efficiency, support policy decisions, and quantify the impact of investment in scientific research on population health and wellbeing. Funding was provided from a consortium of 10 UK Government and Charity Funders led by Medical Research Council (MRC).	England, Scotland, Wales	http://www.swansea.ac.uk/medicine/research/researchthemes/patientpopulationhealthandinformatics/ehealth-and-informatics-research/thefarrinstitutecipher/
European Patients - Smart open Services (epSOS)	A large-scale pilot project on cross-border sharing of personal health data from electronic clinical records, including the sharing of patient summaries and e-prescriptions among selected facilities and professionals within the EU. All participants have agreed to a model for data sharing. The project received funding support from the ICT Policy Support Programme, as part of the Competitiveness and Framework Programme of the European Commission.	Project grew to include 25 European countries	http://www.epsos.eu/
Pharmaco-epidemiological Research on Outcomes of Therapeutics by a European Consortium (PROTECT)	To monitor the benefit-risk of medicines in Europe by developing innovative methods to enhance early detection and assessment of adverse drug reactions from different data sources (clinical trials, spontaneous reporting and observational studies) and to enable the integration and presentation of data on benefits and risks. PROTECT is supported by the Innovative Medicines Initiative.	European countries	http://www.imi-protect.eu/objectives.shtml

Projects leveraging data linkages and electronic health record data for health system performance and quality comparisons

Title	Description	Participating countries	To learn more
European Health Care Outcomes, Performance and Efficiency Project (EuroHOPE)	To evaluate the performance of European health care systems in terms of quality and efficiency by developing performance indicators that measure the use of resources and the health-outcomes of care for uniformly-defined patient groups (disease groups) that cover the entire episode of care, using standardised risk-adjustment procedures. Focus of interest is in variation at the hospital, regional and national levels as well as to explore and reveal reasons behind differences in outcomes and costs. EuroHOPE is supported by the European Union (7th Framework).	Finland, Hungary, Italy, Netherlands, Norway, Scotland, Sweden, and Denmark	http://www.eurohope.info/
European Collaboration for Healthcare Optimisation (ECHO)	To provide detailed performance comparisons between providers and areas within health systems to shed light on performance variations. Variations are presented at the level of individual providers (hospitals or healthcare areas). Through identification of variations, the evidence can be used toward improving equity, quality, safety and efficiency. ECHO is supported by the European Union (7th Framework).	Austria, Denmark, England, Portugal, Slovenia and Spain	http://www.echo-health.eu/
Nordic Quality Measurments in Health Care	To enable Nordic residents, politicians, health care personnel and health authorities to assess and compare the quality of health services across national borders. Another aim is to identify areas where Nordic countries can learn from one atnother to improve the quality of health services for patients.	Denmark, Faroe Islands, Finland, Greenland, Iceland, Norway, Sweden	http://www.norden.org/en/publications/publikationer/2010-572

Projects linking cancer registry data to other data sources to compare and explain cancer survival differences internationally

Title	Description	Participating countries	To learn more
CONCORD-2	An international (worldwide) cancer survival comparison covering ten cancer sites in adults, and childhood leukaemia, with data from over 270 cancer registries in 61 countries on patients diagnosed during the period 1995-2009 or later. The project aims to examine the underlying causes of survival differences, and to derive measures such as the population "cure" fraction, cancer prevalence and the number of avoidable premature deaths, as a basis for informing national and global policy for cancer control. CONCORD is supported by the Union for International Cancer Control.	Over 270 cancer registries in 61 countries	http://www.lshtm.ac.uk/eph/ncde/cancersurvival/research/concord/index.html
Eurocare	Eurocare has been under way for more than 20 years and aims to describe and explain geographical variation in cancer survival across Europe and encourage improvements in cancer care. The fifth and current edition includes data on more than 21 million cancer diagnoses. Eurocare is supported by the Italian Ministry of Health, the Compagnia di San Paolo di Torino and the CARIPLO Foundation.	116 Cancer Registries in 30 European countries	http://www.eurocare.it/Home/tabid/36/Default.aspx
International Cancer Benchmarking Partnership	To understand how and why cancer survival varies between countries. The project has examined cancer survival rates of four cancer types: breast, colorectal, lung and ovarian cancer and has released stage-specific survival results.	Australia, Canada, Denmark, NorwaySweden, and United Kingdom	http://www.cancerresearchuk.org/cancer-info/spotcancerearly/ICBP/
Nordic Occupational Cancer Study	To better understand the role of occupational exposures in the etiology of cancer. A cohort study of 15 million people 30-64 years. Population-based cancer registries are linked to information on occupation from censuses. The project is sponsored by the Nordic Cancer Union and Scientific Council in Sweden.	Denmark, Finland, Iceland, Norway and Sweden	http://ki.se/en/meb/nocca
Survival from six adult cancers in the UK and Republic of Ireland	To describe and compare the one-year and five-year survival for six adult cancers, including cancers of the breast, lung, colon, rectum, melanoma and ovary to investigate the trend in international survival and regional variation in survival estimates.	England, Wales, Scotland, Northern Ireland and Republic of Ireland	

Key features of high-value, privacy-protective health information systems

A key set of OECD countries have emerged from this study as having made possible the development of national health information systems that are capable of providing information to monitor and improve health and health care quality, safety and performance. The development of such powerful health information systems needs to be accomplished within a data governance framework that protects patients' health information privacy. Such frameworks require the involvement of the society in their initiation and evolution so that they reflect societal values. The development of the health information systems and the governance of the data within them require genuine, meaningful consultation with key stakeholders.

The remainder of this report will focus on the essential data governance mechanisms that countries can put into place to reach a high-value and privacy-protective health information system that deserves the public's trust. With appropriate data governance that respects patient's rights to privacy, high-value health information systems can be developed to support the public's right to health and to high-quality, safe and efficient health care.

The Advisory Panel of Experts on Health Information Infrastructure identified the following key features of high-value, privacy-protective health information systems:

1. The health information system supports the monitoring and improvement of health care quality and system performance, as well as research innovations for better health care and outcomes.

The health information system:

a) Is accessible for statistics and research, subject to safeguards specified in the legislative framework.

b) Is developed within a data governance framework that protects health information privacy and reflects societal values regarding rights to privacy and to health.

c) Is developed by establishing information priorities, data collection requirements and data content standards through formal and open consultation with key stakeholders.

d) Includes datasets of patient-level data for complete or representative national patient populations for all key health and social care services and for patient characteristics, behaviours and health outcomes.

e) Includes data from clinical, administrative, laboratory, device and survey sources that can be linked and analysed for approved statistics and research projects.

f) Includes the collection of consistent, patient identifiers for datasets where identification and/or data linkage is in the public interest*.

g) Follows international standards for the coding of terminology and data interoperability.

h) Is routinely audited for information content quality and usability for research and statistics.

i) Enables datasets to be routinely linked for approved on-going monitoring of population health, health care quality and system performance in the public interest*.

j) Enables datasets to be routinely linked for approved research projects in the public interest*.

* The notion of public interest includes: data protection, public health, social protection, the management of health care services, health research and statistics.

Notes

1. For this study, members of the United Kingdom, England, Scotland and Wales, are reported individually due to significant and important differences in their health information system development and governance that are of interest to other OECD countries.
2. CIHI National Prescription Drug Utilization Information System and Canadian Primary Care Sentinel Surveillance Network.
3. The Czech Republic will collect these data through an online survey of community care.

References

Allemani, C. et al. (2013), "Breast Cancer Survival in the US and Europe: A CONCORD High-Resolution Study", *International Journal of Cancer*, Vol. 132, No. 5, pp. 1170-1181.

Cancer Research UK, ICBP Publications, www.cancerresearchuk.org/cancer-info/spotcancerearly/ICBP/icbp-publications/, accessed 14 September 2013.

CIHI – Canadian Institute for Health Information (2013), *Insights and Lessons Learned From the PHC VRS Prototype*, Canadian Institute for Health Information, see www.cihi.ca/CIHI-ext-portal/pdf/internet/LESSONS_PHC_VRS_PROTO_EN, accessed 13 February 2015.

CIHI (2012), *Pathways of Care for People with Stroke in Ontario*, Canadian Institute for Health Information, Ottawa, https://secure.cihi.ca/free_products/Pathways_of_care_aib_en.pdf, accessed 7 August 2014.

Coleman, M.P. et al. (2011), "Cancer Survival in Australia, Canada, Denmark, Norway, Sweden, and the UK, 1995-2007 (the International Cancer Benchmarking Partnership): An Analysis of Population-based Cancer Registry Data", *The Lancet*, Vol. 377, No. 9760, pp. 127-138.

Coloma, P.M. et al. (2012), "Electronic Healthcare Databases for Active Drug Safety Surveillance: Is There Enough Leverage?", *Pharmacoepidemiology and Drug Safety*, Vol. 21, pp. 611-621.

Daňková, Š. et al. (2011), *Analysis of the Recipients of "Care Allowances" as a Potential Clients of Long-Term Care*, Ministry of Employment and Social Affairs.

Denny, J.T. (2010), "Family and Household Formations and Suicide in the United States", *Journal of Marriage and the Family*, Vol. 72, No. 1, pp. 202-213.

European Collaboration for Healthcare Optimisation (ECHO) (2013), www.echo-health.eu/, accessed 14 September 2013.

Farr Institute (2014), www.farrinstitute.org/, accessed 12 August 2014.

FDA – Food and Drug Administration (2013), "Mini-Sentinel", Food and Drug Administration, United States, http://mini-sentinel.org/, accessed 13 September 2013.

Gorina, Y. and E.A. Kramarow (2011), "Identifying Chronic Conditions in Medicare Claims Data: Evaluating the Chronic Condition Data Warehouse Algorithm", *Health Services Research*, Vol. 46, No. 5, pp. 1610-1627.

Groenwold, R.H. et al. (2011), "Balance Measures for Propensity Score Methods: a Clinical Example on Beta-agonist Use and the Risk of Myocardial Infarction", *Pharmacoepidemiol Drug Safety*, Vol. 20, No. 11, pp. 1130-1137.

Häkkinen, U. et al. (2013), "Health Care Performance Comparison Using a Disease-based Approach: The EuroHOPE Project", *Health Policy*, Vol. 112, No. 1, pp. 100–109.

Haklai, Z. et al. (2010), *Health in Israel. Selected data 2010*, Ministry of Health.

Haklai, Z. et al. (2011), "The Mortality Risk Among Persons with Psychiatric Hospitalizations", *Israel Journal of Psychiatry and Related Sciences*, Vol. 48, No. 4.

HSJ – Health Service Journal (2011), "HSJ Best Practice Awards 2011", Health Service Journal, www.hsj.co.uk/journals/2011/11/21/o/c/i/HSJBP2011.pdf, accessed 7 August 2014.

ICES – Institute for Clinical and Evaluative Sciences (2013), *At a Glance: Evidence Guiding Health Care*, Institute for Clinical and Evaluative Sciences, Toronto, www.ices.on.ca/webpage.cfm?site_id=1&org_id=70, accessed 13 September 2013.

Keshavjee, K. et al. (2014), "Getting to Usable EMR Data", *Canadian Family Physician*, Vol. 60, p. 392.

Kumakawa, T. (ed.) (2013), "Strategic Management of Evidence-based Health and Medical Care Policy: How to Use New Digital Big Data in Health Care System", Special issue of the *Journal of the National Institute of Public Health,* Vol. 62.

Linnet, K. et al. (2013), "Primary Non-adherence to Prescribed Medication in General Practice: Lack of Influence of Moderate Increases in Patient Copayment", *Family Practice*, Vol. 30, No. 1, pp. 69-75.

Livingstone, S.J. et al. (2012), "Risk of Cardiovascular Disease and Total Mortality in Adults with Type 1 Diabetes: Scottish Registry Linkage Study", *PLoS Med*, Vol. 9, No. 10.

Maringe, C. et al. (2013), "Stage at Diagnosis and Colorectal Cancer Survival in Six High-income Countries: A Population-based Study of Patients Diagnosed During 2000-7", *Acta Oncologica*, Vol. 52, No. 5, pp. 919-932.

Maringe, C. et al (2012), "Stage at Diagnosis and Ovarian Cancer Survival: Evidence from the International Cancer Benchmarking Partnership", *Gynecologic Oncology*, Vol. 127, No. 1, pp. 75–82.

Merchante, J.M. et al. (2010), "Analysis of the Agency for Healthcare Research and Quality Patient Safety Indicators in Public Hospitals of the Autonomous Region of Madrid (Spain)", *Medicina Clínica (Barcelona)*, Vol. 135, Suppl. 1, pp. 3-11.

Næss, Ø., B.H. Strand and G. Davey Smith (2007), "Childhood and Adulthood Socioeconomic Position Across 20 Causes of Death: A Prospective Cohort Study of 800 000 Norwegian Men and Women", *Journal of Epidemiology & Community Health*, Vol. 61, pp. 1004-1009.

NHS – National Health Service (2014), www.nhs.uk/NHSEngland/thenhs/records/healthrecords/Pages/care-data.aspx, accessed 11 August 2014.

NHS – National Health Service, England (2013), "Care Episodes Statistics: Technical Specifications of the GP Extract", National Health Service, May, www.england.nhs.uk/wp-content/uploads/2013/08/cd-ces-tech-spec.pdf, accessed 31 July 2015.

OECD (2015), *Data Driven Innovation for Growth and Well-Being*, OECD Publishing, Paris, forthcoming.

OECD (2013a), *Strengthening Health Information Infrastructure for Health Care Quality Governance: Good Practices, New Opportunities and Data Privacy Protection Challenges*, OECD Health Policy Studies, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264193505-en.

OECD (2013b), *Health at a Glance 2013*: OECD Indicators, OECD Publishing, Paris, http://dx.doi.org/10.1787/health_glance-2013-en.

Okamoto, E. et al. (2013), "Evaluation of the Health Check up and Guidance Program through Linkage of Health Insurance Claims", *Journal of the National Institute of Public Health,* Vol. 62, No. 1.

Pukkala, E. et al. (2009), "Occupation and Cancer – Follow-up of 15 Million People in Five Nordic Countries", *Acta Oncologica*, Vol. 48, No. 5, pp. 646-790.

Sanni Ali, M. et al. (2013), "Time-Dependent Propensity Score and Collider-Stratification Bias: An Example of Beta2-Agonist use and the Risk of Coronary Heart Disease", *European Journal of Epidemiology*, Vol. 28, No. 4, pp. 291-299.

Sant, M. et al. and EUROCARE Working Group (2009), "EUROCARE-4. Survival of Cancer Patients Diagnosed in 1995-1999. Results and Commentary", *European Journal of Cancer*, Vol. 45, No. 6, pp. 931-991.

Spillane, S. et al. (2013), "A Cohort Study of Metformin Exposure and Survival in Patients with Stage I-III Colorectal Cancer", *Cancer Epidemiology, Biomarkers and Prevention*, Vol. 22, No. 8, pp. 1364–1373.

Spoerri, A. et al. (2010), "The Swiss National Cohort: A Unique Database for National and International Researchers", *International Journal of Public Health*, Vol. 55, No. 4, pp. 239-242.

Socialstyrelsen (2013a), *Quality and Efficiency in Swedish Health Care: Regional Comparisons, 2012*, Stockholm.

Socialstyrelsen (2013b), *Quality and Efficiency of Stroke Care in Sweden*, Stockholm.

Walters, S. et al. (2013a), "Comparability of Stage Data in Cancer Registries in Six Countries: Lessons from the International Cancer Benchmarking Partnership", *International Journal of Cancer*, Vol. 132, No. 3, pp. 676–685.

Walters, S. et al. (2013b), "Breast Cancer Survival and Stage at Diagnosis in Australia, Canada, Denmark, Norway, Sweden and the UK, 2000-2007": A Population-based Study", *British Journal of Cancer,* Vol. 108, pp. 1195–1208.

Walters, S. et al. (2013c), "Lung Cancer Survival and Stage at Diagnosis in Australia, Canada, Denmark, Norway, Sweden and the United Kingdom: A Population-based Study, 2004-2007 (2013)", *Thorax,* Vol. 68, No. 6, pp. 551-564.

Webster, G., P. Sullivan-Taylor and T. Flanagan (2011), "Maximizing EMR Benefits Through Data Standards & CIHI's Primary Health Care Voluntary Reporting System", *Healthcare Information Management & Communication Canada (HIM&CC)*, 2nd Quarter.

Williamson, T. et al. (2014), "CPCSSN's Role in Improving Pharmacovigilance", *Canadian Famoly Physician*, Vol. 60, p. 678.

Chapter 3

The legislative framework governing personal health data

This chapter provides an overview of the features of OECD countries' legislative frameworks and their application in practice and sets out the key factors within legislative frameworks that support privacy-protective health and health care monitoring and research.

It concludes by identifying aspects of legislative frameworks that protect patient privacy while enabling data to be used for research and statistics.

Highlights

Throughout the OECD, the legal framework for the protection of personal data recognises health data as sensitive data that require a high level of protection. Most countries have more than one national legislation that governs aspects of health information privacy protection and In some federated countries there are also provincial or state laws governing personal health data.

Countries were asked about a set of data accessibility factors that are directly linked to legislative frameworks and their interpretation in practice. Overall, data sharing and accessibility is greatest in New Zealand, Sweden and the United Kingdom. Nine countries, however, do not permit the sharing of personal data among national holders of all or most key datasets. Further, even after data have been de-identified, two countries have no mechanism to permit academic researchers to analyse it; seven countries have no mechanism for applicants from the commercial sector to analyse it, even if their work has a public benefit; and five countries have no mechanism for applicants from a foreign country to analyse it, even if the project has public benefits nationally and internationally.

Whether enabled by a broad and prospective patient consent, exemption to patient consent requirements or legal authorisation, the legislative framework should enable countries to approve privacy-protective uses of personal health data that are in the public interest, including extraction of data from electronic clinical records and dataset linkages. Legislative frameworks need careful development to ensure that health information privacy protections are consistent for all forms of personal health data and that there are no forms of personal data that fall outside of legal protection; and to provide mechanisms for citizens to express their choices regarding uses of their personal health data for statistics and research that are fair and practicable.

The legislative framework is the foundation upon which a country's health information infrastructure may be developed and citizens' rights to health and to privacy codified. While health sector specific legislation is often developed in collaboration with national health authorities, legislation regarding the protection of the privacy of individuals and their personal information may be developed by justice ministries or other areas of government. As a result, it is of fundamental importance to health authorities planning to strengthen their health information infrastructure, that there is open dialogue across government regarding the legislative framework necessary to support maximising societal benefits from health data while minimising societal risks from data uses.

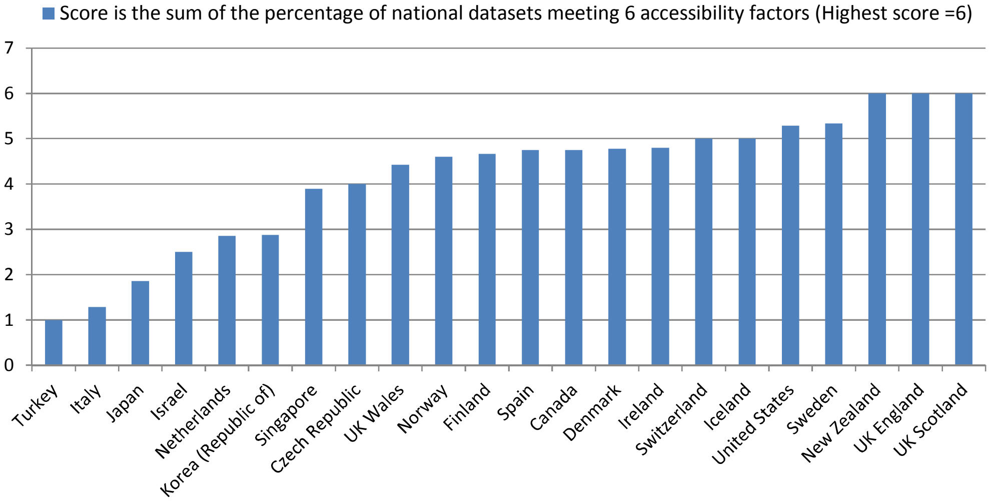
This chapter introduces differences in data accessibility among OECD countries; discusses the legislative frameworks for the protection of personal health data; explains how legislative frameworks permit or restrict research and statistical uses of data; discusses current approaches to and discussions about patient consent; discusses legislative permissions and restrictions regarding the sharing of identifiable and de-identified microdata; explains current approaches to requests for access to data from foreign countries; and discusses data sharing challenges among public authorities and between national authorities and providers of health care services. It concludes by identifying aspects of legislative frameworks that protect patient privacy while enabling data to be used for research and statistics.

Data accessibility across OECD countries

In the 2013 OECD country survey, respondents were asked about a set of key data accessibility factors that are directly linked to legislative frameworks and their interpretation in practice. These factors include whether or not identifiable national personal health data are ever shared among data custodians or government entities and whether personal health data, after de-identification, can be approved for access by applicants from different sectors of society and by foreign applicants.

Findings are summarised in Figure 3.1. Overall data sharing and accessibility is greatest in New Zealand and the United Kingdom. Countries with the lowest sharing and accessibility of health data are Turkey, Italy and Japan.

Figure 3.1. Sharing and accessibility of health data for approved statistical and research uses



Source: Author's own calculations based on the results of this study.

Results for each of the six accessibility factors are reported by country in Table 3.1. Five countries reported that none of the key national health datasets is ever shared in an identifiable format with another data custodian or government entity. As is explored in more detail in other chapters of this report, countries that prohibit the sharing of identifiable data among government authorities may still be able to develop data about health care pathways and outcomes. They do so either because many key datasets are in the custody of a single organisation (see Chapter 2) or, alternatively, because they have good co-operation among different government entities, and each entity agrees to encrypt identifying variables using the same algorithm, enabling the linkage of de-identified data (see Chapter 7).

Virtually all countries reported that analysts from a government authority could apply for and be approved access to the majority of key national de-identified micro datasets. A micro dataset contains records for patients or persons. Only Italy restricts government authorities from accessing de-identified microdata for the majority of national datasets.

Israel and Turkey restrict non-profit and university based researchers from access to the majority of national de-identified health micro datasets.

A small group of countries has excluded health care providers from access to any of the key national de-identified health micro datasets (Italy, Japan, Korea, and Turkey). Israel provides access for health care providers to half of its key national de-identified micro datasets.

A larger group of countries do not permit analysts from for-profit businesses to be approved access to de-identified microdata for any or virtually all of the key national datasets: Czech Republic, Israel, Italy, Japan, Korea, Singapore and Turkey. Countries that permit commercial organisations to be approved access to data, restrict the use of the data to scientific research and statistics that are in the public interest.

In several countries, access to de-identified microdata are also restricted for university or non-profit and government applicants from a foreign country. No access is allowed in Israel, Italy, Japan, Korea and Turkey. Further access to less than half of national de-identified datasets is permitted in Ireland and access is limited to only the de-identified cancer registry in Singapore.

Table 3.1. Proportion of key national personal health datasets meeting six data accessibility factors

	Identifiable data is shared with other data custodian or government entities	Government analysts may be approved access to de-identified data	University and non-profit researchers may be approved access to de-identified data	Health care providers may be approved access to de-identified data	For-profit businesses may be approved access to de-identified data	Foreign government, university or non-profit researchers may be approved access to de-identified data
Canada	75%	88%	88%	88%	75%	63%
Czech Republic	0%	100%	100%	100%	0%	100%
Denmark	78%	89%	89%	89%	78%	56%
Finland	78%	78%	78%	78%	78%	78%
Iceland	0%	100%	100%	100%	100%	100%
Ireland	80%	100%	100%	100%	60%	40%
Israel	67%	100%	33%	50%	0%	0%
Italy	14%	29%	86%	0%	0%	0%
Japan	0%	86%	100%	0%	0%	0%
Korea (Rep. of)	100%	100%	88%	0%	0%	0%
Netherlands	14%	71%	71%	71%	ns	57%
New Zealand	100%	100%	100%	100%	100%	100%
Norway	57%	100%	60%	100%	40%	100%
Singapore	75%	100%	100%	100%	0%	13%
Spain	75%	100%	75%	75%	75%	75%
Sweden	89%	89%	89%	89%	89%	89%
Switzerland	0%	100%	100%	100%	100%	100%
Turkey	0%	100%	0%	0%	0%	0%
United States	29%	100%	100%	100%	100%	100%
UK England	100%	100%	100%	100%	100%	100%
UK Scotland	100%	100%	100%	100%	100%	100%
UK Wales	14%	86%	86%	86%	86%	86%

ns: Not stated.

Source: Author's own calculations based on the results of this study.

Legislative frameworks for the protection of privacy

Legislations and privacy policies have been influenced by the 1980 publication of the OECD privacy guidelines and these guidelines are recognised as representing "the international consensus on privacy standards and providing guidance on the collection of personal information in any medium" (OECD, 2009). The OECD guidelines emphasize that data collections are respectful of the protection of personal privacy when they follow eight guiding principles (Box 3.1). In 2013, the OECD revised these guidelines; however, the eight guiding principles remain relevant and were unchanged (OECD, 2013).

These principles were subsequently reflected in the 1995 *Data Protection Directive* of the European Union (95-46-EC) that regulates the processing of personal information. In the European Union, a directive is a legal act that is required as a result of an EU treaty. Directives are binding for member states and each state is required to incorporate the directive into law within the time period specified in the directive.

Following the directive, European countries have implemented specific legislation relating to the protection of the privacy of personal information that complies with EU regulatory requirements. All of the European countries participating in this study report the existence of data protection legislation and an oversight body responsible for guidance and monitoring of this legislation in the form of a privacy or data protection office at the national level. While providing a unifying framework, the directive left considerable freedom to countries regarding whether to apply, restrict or extend the rules on processing sensitive data. In 2012, the European Union published a proposal for a new data protection regulation (European Commission, 2012).

This new regulation is subject to amendment as it proceeds through the parliamentary process toward final ratification. There is considerable uncertainty regarding whether this regulation will provide clarity regarding the use of personal health data for research and statistics. Current indications are that the regulation will place a higher emphasis upon specific informed consent than did the 1995 directive and will allow for derogation from this principle only by law and for matters of exceptionally high public interest (Di Iorio, 2013).

There is a key national legislation that speaks to the protection of health information privacy in all countries participating in this study (Table 3.2). Most countries have more than one national legislation that governs aspects of health data privacy protections. In many countries there is both general data privacy legislation applying to all personal data and health-sector specific legislation providing greater clarity regarding the collection and use of personal health data. In some federated countries, including Canada and the United States, there are also provincial or state laws governing personal health data.

Box 3.1. Guiding principles for the protection of privacy and the transborder flow of personal data

The OECD guidelines for the protection of privacy and the transborder flow of personal data outline eight guiding principles for national application:

1. Collection limitation principle	There should be limits to the collection of personal data and any such data should be obtained by lawful and fair means and, where appropriate, with the knowledge or consent of the data subject.
2. Data quality principle	Personal data should be relevant to the purposes for which they are to be used and, to the extent necessary for those purposes, should be accurate, complete and kept up-to-date.
3. Purpose specification principle	The purposes for which personal data are collected should be specified not later than at the time of data collection and the subsequent use limited to the fulfillment of those purposes or such others as are not incompatible with those purposes and as are specified on each occasion of change of purpose.
4. Use limitation principle	Personal data should not be disclosed, made available or otherwise used for purposes other than those specified in accordance with Paragraph 9 except a) With the consent of the data subject; or b) By the authority of law.
5. Security safeguards principle	Personal data should be protected by reasonable security safeguards against such risks as loss or unauthorised access, destruction, use, modification or disclosure of data.
6. Openness principle	There should be a general policy of openness about developments, practices and policies with respect to personal data. Means should be readily available of establishing the existence and nature of personal data, and the main purposes of their use, as well as the identity and usual residence of the data controller.
7. Individual participation principle	An individual should have the right: a) To obtain from a data controller, or otherwise, confirmation of whether or not the data controller has data relating to him; b) To have communicated to him, data relating to him within a reasonable time; at a charge, if any, that is not excessive; in a reasonable manner; and in a form that is readily intelligible to them; c) To be given reasons if a request made under subparagraphs(a) and (b) is denied, and to be able to challenge such denial; and d) To challenge data relating to them and, if the challenge is successful to have the data erased, rectified, completed or amended.
8. Accountability principle	A data controller should be accountable for complying with measures which give effect to the principles stated above.

Source: OECD (2009), *OECD Policies for Information Security and Privacy*. Principles re-confirmed in OECD (2013), *The OECD Privacy Framework*.

Table 3.2. Countries reporting a national law or regulation that speaks to the protection of health information privacy and/or to the protection and use of electronic clinical records

Canada	Each jurisdiction in Canada (Federal, Provincial and Territorial) has its own privacy legislation that governs the collection, use and disclosure of personal information. In addition, some provinces have specific health information privacy legislation and/or legislation that specifically address the health information in EHRs. Consequently, health information will be afforded protection under a variety of privacy and health information protection laws in Canada. At the national level, there are two federal acts that may apply: The Privacy Act and the Personal Information Protection and Electronic Documents Act. The website of the Federal Privacy Commissioner is a good source of more information on these pieces of legislation (http://www.priv.gc.ca/leg_c/leg_c_p_e.asp).
Czech Republic	The main general privacy legislation is act no.101/2000 Coll. on The protection of personal data (see http://www.uoou.cz/uoou.aspx?menu=4&submenu=5&lang=en). The privacy of health sector data are protected within act no.372/2011 Coll. on Health care services and act no. 89/1995 Coll. on State statistical service.
Denmark	The Act on Processing of Personal Data (Act No. 429 of 31 May 2000) entered into force on 1 July 2000. The act implements Directive 95/46/EC on the protection of individuals with regard to the processing of personal data and on the free movement of such data. (https://www.retsinformation.dk/Forms/r0710.aspx?id=828)
Finland	The Personal Data Act came into force in 1999 and was introduced as a result of the EU Data Protection Directive. This Act governs all forms of personal data. The Act on the Openness of Government Activities came into force in 1999 and refers to public institutions. In Finland, most health data is in the custody of public institutions, including national institutions such as THL and regional governments and public clinics. This Act has provisions regarding the use of classified and sensitive public data, such as personal health data. There is an Act on the Status and Rights of Patients which provides a broad framework governing clinical research including research involving personal health data. There is a STAKES Act on Social Registers that provides the enabling legislative authority for the development of health data registries within THL. There is a new Act governing the national electronic health record system including e-prescriptions. This Act contains provisions governing the development and use of a national data archive of electronic health records. There is a new Biobank Act enacted in September 2013 with provisions related to data collection, use and access. There is also a Medical Research Act with provisions regarding patient consent.
Iceland	Act no.77/2000 on The Protection of Privacy as regards the Processing of Personal Data; Act no.55/2009 Health Records Act
Ireland	Data Protection Acts
Israel	Privacy Protection Law, 1981; Patients Rights Law, 1996
Italy	Data protection regulation, according to Data Protection Code - Legislative Decree no. 196/2003
Japan	Privacy Protection Act and legislation governing health providers and health insurance
Korea	Personal Information Protection Act, National Health Insurance Act, Medical Care Assistance Act, Act on Support for Persons Eligible for Veterans, Medical Service Act, and Pharmaceutical Affairs Act. The Personal Information Protection Act was introduced on October 30, 2011 and provides the general legal framework for the protection of privacy including use limitations. Other legislations stipulate the exceptions to the general rules provided within the Personal Information Protection Act. This includes the provisions within the Act on the Protection of Personal Information mandated by Public Institutions. There are sometimes conflicts between the applicable laws. If there is a need in the public interest for a use of personal health data and the use is not permitted under PIPA, then the use is either authorised under legislation or a new legislation is introduced to enable the data use.
Netherlands	There are several laws.
New Zealand	Privacy Act 1993 (including Health Information Privacy Code 1994 (HIPC)); Health Act 1956. The Privacy Act applies to all personal data and applies to public and private data holders. The Health Information Privacy Code (HIPC) applies to the health sector including public institutions, public hospitals, private health practitioners, and insurers. The code provides rules regarding personal health data including uses, disclosure, uses of health numbers, data retention and data security requirements.
Norway	Law on health records for processing data. Law on the processing of personal data. Law on health personnel and practice regulations.
Singapore	Private Hospitals & Medical Clinics (PHMC) Act, Personal Data Protection Act, National Registry of Diseases Act, Infectious Diseases Act, Termination of Pregnancy Act.
Spain	Ley Orgánica de Protección de Datos and derived regulations
Sweden	There are two main laws related to the protection of health information privacy: The Public Access to Information and Secrecy Act and the Personal Data Act. The Public Access to Information and Secrecy Act applies to public authorities and also to certain private companies founded by a public authority and undertaking work that would have been undertaken by a public authority. Such private companies are under a public authority and are under this Act. The Personal Data Act applies to any holder of personal data in Sweden (whether public or private) and protects against violations of individuals rights to privacy. Public authorities, such as the National Board of Health and Welfare (NBHW) must review the requirements of both of these Acts when considering the approval of projects involving personal health data.
Switzerland	Health policy is a domain of the cantons; they have individual laws on privacy, some explicitly for EHR.
Turkey	Patient's Rights Directive
United States	Health Insurance Portability and Accountability Act (HIPAA), which requires the protection and confidential handling of protected health information; the Federal Privacy Act (1974); and state-level legislations whose protections may supersede federal law if they provide for stronger privacy protections.
United Kingdom	UK Data Protection Act 1998 and the Common Law Duty of Care govern the use of personal health data. Common-law develops over time on the basis of legal precedents (past legal decisions).

Source: OECD Health Care Quality Indicators Country Survey, 2013/14.

Personal health data can have inconsistent legislative protection

In general, national legislations apply to all processing of personal health data and therefore cover all personal health data in the country. There are, however, gaps in some national legislative frameworks that create inconsistencies in privacy protection or result in some personal health data falling through the cracks and having no legislative protection.

There's no general data privacy protection legislation at the national level in Canada. There is a patchwork involving federal, provincial/territorial and municipal laws, as well as acts specific to health information and health professionals. Each of the 13 provinces and territories has specific legislation related to the privacy of personal information, some have specific legislation on the protection of health information in place and others have health sector-specific privacy legislation pending. At the national level, Canada has a *Privacy Act*, which covers the personal information-handling practices of federal government departments and agencies, and the *Personal Information Protection and Electronic Documents Act* (PIPEDA), the federal private-sector privacy law. PIPEDA sets out how private-sector organisations collect, use or disclose personal information in the course of commercial activities across Canada. It also applies to federally-regulated works, undertakings, or businesses. Physicians in Canada are subject to PIPEDA because they are engaged in commercial activities. PIPEDA will not apply, however, where the organisation operates within a jurisdiction that has legislation that has been deemed substantially similar to the PIPEDA, unless the personal information crosses provincial or national borders. Jurisdictional privacy legislations are based on the same privacy protection principles and aim to be substantially similar to PIPEDA so that the jurisdictional legislation can apply to all personal information. It is unlikely that there are personal health data excluded from some legislative protection in Canada.

In New Zealand, the *Health Information Privacy Code* (HIPC) applies to the health sector including public institutions, public hospitals, private health practitioners, and insurers but does not apply to health data held by private companies. The HIPC provides rules regarding personal health data including uses, disclosure, uses of health numbers, data retention and data security requirements. Health data held by private companies does not fall under the HIPC. However, the general rules set in the national *Privacy Act* and in the code are consistent and the *Privacy Act* applies to all personal data and applies to public and private data holders.

In the United States, the main national health data privacy law is the *Health Insurance Portability and Accountability Act* (HIPAA). There are regulations under HIPAA that have extensive provisions for data privacy, security, and breach notification. Nearly all personal health information in the health care treatment and payment systems are subject to HIPAA. Health data held outside of treatment and payment systems may have little legislative protection.

There are laws at the state level in the United States that will take precedence over HIPAA wherever the state law offers a higher level of data protection. The degree of fragmentation in law among US states has been reducing, however, as many states have revised their laws to have more in common with HIPAA. Nonetheless, national research is complex in the United States, as the state legal requirements must be met. Unless protected by state law, there are numerous holders of personal health data exempt from any privacy protection laws, particularly private-sector and university-based holders of personal health data (see Box 3.2).

In Singapore, the Personal Data Protection Act (PDPA) governs all identifiable (personal) data including personal health data. Health sector specific legislation applies to all licensed health care institutions and requires purpose-specific use of health data and obliges health care institutions to protect personal health data. Non-licenced entities are not governed by the health sector specific legislation but are still subject to PDPA requirements.

Box 3.2. US national health information privacy law offers strong protection to a narrow set of data custodians and data types

Many organisations in custody of personal health data are not included under the Health Insurance Portability and Accountability Act (HIPAA), such as the National Institutes of Health (NIH). The Federal Privacy Act of 1974 applies to all federal departments and agencies and the NIH is under this legislation. Authorising legislation for federal departments and agencies can also regulate the protection of data privacy but with varying rules for each. The National Center for Health Statistics, for example, has its own authorising legislation with specific provisions for the protection of data confidentiality. Schools and colleges with personal health data are not covered under HIPAA but are covered under education-sector privacy protection law.

There are, however, numerous custodians of personal health data that do not fall under any legal protection in the United States. Examples include non-medical health service providers (gyms, nutrition counsellors, etc.), private labs and genetic testing services, patient reported data from marketing surveys, fitness apps and social media, banks, life insurance providers and health care providers that don't accept insurance, such as employer clinics. Identifiable data are often sold among unregulated data holders, particularly for marketing purposes. Importantly, researchers who don't work for health insurers or health care providers or national government departments and agencies, do not fall under any legislation to protect privacy. This would include most researchers within universities and colleges and non-profit and for-profit research institutes. Further, data that are protected under HIPAA lose this legal protection when it is transferred from a covered entity to an uncovered researcher. Instead, researchers in the United States may fall under the common rule.

The common rule provides protection to human subjects of medical research. The common rule requires that any project involving the use of personal health data without consent be approved by a research ethics board before it can be conducted. All federal departments and agencies follow the common rule including the NIH. The rule applies to research supported by federal grant funding or where research results may require federal regulatory approval (such as clinical trials of drugs that would require federal government approval). The common rule does not bind researchers to data security requirements nor does it hold researchers accountable for poor data security.

An advisory committee to the Department of Health and Human Services in the United States reported on the gaps in health data protection in the United States, however, there is no concerted effort toward a solution.

National health datasets contain sensitive personal information

Throughout the OECD, the legal framework for the protection of personal data recognises health data as sensitive data and therefore requiring a high level of protection. There are particular variables within national health datasets, however, that may be considered to be of even higher sensitivity than other variables. Table 3.3 presents the types of variables that countries identified as being among the most sensitive within their national datasets. Variables that lead to the direct identification of individuals are highly sensitive, as are particular health conditions that may carry additional social stigma, namely self-harm, mental health conditions, sexually transmitted infections including HIV, substance use and treatment, sexual health, abortion, child abuse and homicide.

Table 3.3. Variables considered as being among the most sensitive within national health datasets

Health care data (hospital and community)	Prescription medicines data	Cancer, diabetes and CVD registry data	Mortality data	Formal long-term care data
Direct identifiers	Direct identifiers	Direct identifiers	Direct identifiers	Direct identifiers
Residential area	Diagnosis	Residential area	Residential area	Aggressive behaviours
Ethnic group/nationality	All Rx*	Ethnic group/nationality	Ethnic group/nationality	Self harm
Diagnosis	Rx* for mental health	Diagnosis	Cause of death	Mental health diagnosis
Surgical procedure	Rx* for HIV	Cancer stage	Cause of death is homicide	Activity limitations
Diagnosis of self-harm	Rx* for drug abuse	HIV status	Cause of death is self-harm	Diagnosis
Diagnosis of HIV	Rx* for contraception	illicit drug use	Cause of death is HIV/AIDS	Rx
Diagnosis of STI		Cause and date of death	Date of death	
Diagnosis of mental health disorder			Cause of death is drug overdose	
Diagnosis of substance misuse				
Diagnosis of child				
Illicit drug use				
Open abortion (indused, spontaneous or both)				

Rx: Prescribed medicines.

Source: Author's own calculations based on the results of this study.

In a small number of countries there have been legislations or practices introduced for the protection of certain topics of personal health data that have been deemed as more sensitive. For example, in Israel there are specific pieces of legislation for particular types of health/medical information that have been determined to be more sensitive than other personal medical information. Examples include the *Genetic Information Law*; and laws specific to abortion services; fertility treatments; and psychiatric health services. In the United States there are laws to protect certain vulnerable groups from discrimination including discrimination related to uses of their personal health information. This includes the *Americans with Disabilities Act* that relates to limiting the use of health information for discrimination and the *Genetic Information Non-Discrimination Act* which controls the use of genetic information by employers.

In Iceland, the unique person identifier is not captured in the electronic datasets covering abortions and sterilisations. In Canada, certain data may have additional restrictions by policy. For example, data concerning abortions are supressed by the Canadian Institute for Health Information when record-level data are provided to approved third parties. In Denmark, there are particular databases held by the Statens Serum

Institute (SSI) that are not stored with patient identifying numbers and are not available for use in data linkage projects. This includes both the HIV and drug abuse registries.

Legislation may permit the secondary analysis of personal health data in cases where patient consent is not possible or practicable

The secondary analysis of personal health data is typically permitted in countries with the consent of the data subject or when the analysis has been legally authorised. An important difference among countries is in whether or not the national legislation governing data privacy protection has recognised statistics and research as potential areas where an exemption to patient consent requirements could be granted. In these countries, an exemption can be granted for a proposed secondary use of personal health data that are in the public interest. In other countries where such exemptions are not legally permitted under the general data protection legislation, the general law may allow for a legally-authorised exemption to patient consent requirements. In these countries, health-sector specific legislation may be introduced to clarify permitted uses of personal health datasets for statistics and research in the public interest.

In New Zealand, the *Health Information Privacy Code* (HIPC) permits secondary data analysis with consent or authorisation of data subjects (or their representative) and for statistical purposes or for research purposes if the data will be anonymised and could not reasonably be expected to identify the individuals concerned.

In the United Kingdom, the *Data Protection Act* does not require consent. Under the terms of this law, consent is one of a number of conditions that could be relied on to support the processing of personal data. The common law duty of confidence does require consent to be obtained unless the processing is for the direct care of an individual. There is a tension between the two approaches. The national health minister has the power to set aside this common law duty in exceptional cases where it is not realistic to expect to obtain consent and where the processing is in the public interest and is under controls to protect the privacy of data subjects. Where there is an exemption to requirements for explicit consent, the preferred policy in the United Kingdom is to offer an opt-out. Therefore the policy regarding health data in the public sector is to allow persons to object to having their data contribute to statistics and research uses. If they opt-out, then their data will not be used. When people lack confidence that there are adequately strong controls to protect their data when they are used, then they may either withdraw their consent or opt-out. It is not sufficient to only have strong controls; the controls must be transparent to the public and communicated to the public. The United Kingdom has a layered approach to protection of the data subject's privacy. On one side are strong legal protections and on the other side are strong information governance practices that restrict data access.

In Sweden, the law allows exemptions to patient consent requirements for direct health care and for research and statistics. As a result, Sweden has established research ethics boards which can approve projects involving the processing of personal health data both with and without consent. Also, national authorities can process personal health data for statistical purposes without consent. The National Board of Health and Welfare (NBHW) has its own authorising legislation. Existing law enables the NBHW to create a new registry or database involving personal health data; however, in some cases it is necessary to amend a law or to introduce a new law. For example, the NBHW would like to extend the Prescribed Drug Register to include medicines provided to patients in hospital. Existing laws were reviewed by the NBHW and a determination was made that this extension of the registry was not legally authorised. As a result, the law governing this register must be amended.

In Denmark, law permits the Statens Serum Institute (SSI) to create new registries involving only health sector data. The approval of the Data Protection Authority is required but it is not necessary to implement a new legislation to authorise the new registry. To create a new registry requiring data from the health sector and from other sectors, such as education or labour, however, would require a new legislation to be introduced. Within existing legislation, it is possible for an external researcher to apply for a one-time project involving the linkage of data within the health sector and between the health and other sectors for a specific project. The project would require the approval of the Data Protection Authority.

In Iceland, personal health data may be processed only if the data subject has consented to the processing, if the linkage is a legal obligation of the data custodian, if the linkage is of vital public interest, or if the linkage protects the fundamental rights of the data subject. The *Medical Director of Health and Public Health Act* provides the Directorate of Health with the legal mandate to access personal health data without patient consent in order to develop national registries to inform about population health and to monitor access to and quality of health care services. The law allows the registries to be used for health services planning and scientific research. The law also permits the Directorate of Health to initiate ad hoc data collection to fulfil its mandate. However, a revision to the law would be required when the Directorate wants to add a new data collection on an on-going basis. The law allows for secondary analysis of the data provided that the data are not identifiable and cannot be traced to any individual, living or deceased.

In the United States, HIPAA has broad provisions regarding the use of data without consent for public health. Use or disclosure for research is allowed provided the use of the data has been first approved by a research ethics board. Where both HIPAA and a state law may apply to data, however, the state law will take precedence if it is stronger than HIPAA. So where HIPAA may allow a data disclosure without consent, and a state law may require consent, the state law would apply. As a result, it is difficult to determine the legal requirements for data protection that apply for uses of health data that involve multiple states. Challenges in accommodating varying state legal requirements while developing national datasets were reported to the OECD by the AHRQ (Agency for Healthcare Research and Quality).

As a statistical agency, the United States National Center for Health Statistics (NCHS) is governed by the *Public Health Service Act Section 308(d),* the *Confidential Information Protection and Statistical Efficiency Act* and the *Privacy Act*. By law, confidential data collected by the NCHS, as well as the fact of participation in an NCHS survey, are protected from public disclosure.

In the Czech Republic there are specific exemptions in law for certain communicable diseases within the *Health Services Act*. All hospitalised patients sign an informed consent permitting their personal health data to be used according to the law. The *Health Services Act* gives the list of registries that are authorised. If there is a need for a new registry, then the law would have to be revised. In the Czech Republic, data collected by the NHIS is authorised by law to take place without informed consent. Clinical research studies are required to be conducted with informed consent. For linkages between health or clinical data and genetic data, opinion is that consent would be necessary. It would, however, depend on the purpose of the linkage. If the purpose was a clinical research study, then consent should be obtained. As a result, there is legal authorisation to process data that has been put into place. Without legal authorisation, a data controller has to be able to prove they have obtained consent for processing.

In Israel, health data that are identifiable are treated under law as confidential and cannot be shared with a third party. The only exceptions are if patients have given their consent to data sharing or where there is a legal obligation to share the data. Under certain laws, the state requests access to identifiable data. For example, under law, hospitals must report identifiable data to the state. Consent requirements in Israel extend to deceased persons. When there is processing of data related to deceased patients, consent is required from family members.

Personal health data in Finland may be used for secondary purposes where this use has been authorised by law. The use of personal health data for secondary purposes can be approved without informed consent, as the data collection for health and social welfare registers is regulated with specific legislation. The *Personal Data Act* recognises that personal health data may be legally authorised to be used for scientific research and statistics. Personal health data may not be used in decision making about a single registered person.

In Korea, secondary analysis is possible with the consent of the data subject or to fulfil an obligation imposed by or under any act or subordinate statute.

Box 3.3. Controlling all data disclosures by consent: Lessons learned from Maine, United States

In 1998, the state of Maine in the United States introduced a privacy protection law that required written consent to disclose personal health data with few exceptions (Gellman, 2007). Written consent was required for all routine sharing of patient information among providers treating the same patient; for any disclosures to family members other than information about presence and general health condition during an emergency; and for payments.

The introduction of the law was immediately followed by a strong expression of public dissatisfaction. Public objections to the law related to its restriction of disclosures to family members, the clergy, other physicians and the press. The law did permit disclosures for statistical and research purposes, and such disclosures were not the source of public discontent.

A lesson learned in Maine was that disclosures only with consent are not necessarily what the public wants or expects. Instead, what is needed are practical ways for individuals to express their wishes regarding uses of their personal health data that do not impede their expectations for a workable health care system.

The law was revised in 1999 to allow health care practitioners much more discretion to make disclosures without patient consent, including disclosures for treatment, payment activities and health care operations. The changes also made it easier to provide consent by adding oral consent as a new category and by allowing family members to authorise disclosures.

Protection of the privacy of health care providers

The legal framework in New Zealand, Spain and the Czech Republic protects health care providers' privacy and, as a result, it is not possible to report statistics at the provider level without the provider's consent. This restricts the development of performance indicators and the use of such indicators in pay for performance initiatives.

In Israel, there is a trend toward increasing transparency in the reporting of provider-level quality indicators that is meeting resistance from HMOs. There have been indicators published identifying hospitals and others reported with the hospital's identity concealed.

There have been court challenges requiring the ministry to reveal the identities of hospitals in published indicators.

Consent to uses of data in the future that cannot be specified today

The future uses of health data collected today depend upon decisions taken regarding conditions where there may be exemptions to consent requirements and on whether or not the legal framework would permit asking individuals today to consent to uses of their data in the future that cannot be specified in a detailed way.

Informed consent has become the pillar for protecting individual’s autonomy where research involves human subjects. Informed consent requirements in legislation build from professional codes of practice. Informed consent presumes the ability to indicate clearly to a participant the use and the purpose of a particular research activity. This is feasible for a purpose-specific study, such as an invitation to patients to participate in a clinical trial or a survey.

The requirement to obtain patient consent presents significant challenges, however, for health and health care monitoring and research involving large population and patient databases. These databases are collected for other purposes, such as administering the health system or providing clinical care and represent hundreds of thousands to millions of persons.

If patient consent is required to use these very large databases in the future for statistical or research purposes in the public interest, then consent could be either broad or specific. If the consent is study-specific, then obtaining consent requires contacting the original data subjects. Given the datasets are very large and historical; the exercise will generate useable data that is biased toward non-movers and healthier/younger patients, which can compromise the validity and the utility of the findings. Further, attempting to reach large cohorts can be impractical and requires considerable financial resources. Such an approach in an environment where there is also an active programme of data use would not be long tolerated by the public, particularly since individuals appearing in health care databases are ill and often elderly. A case study in Maine is instructive in this respect (see Box 3.3).

If the consent is broad, it does not imply that there is no further governance of data uses. As will be discussed in later chapters, decisions about the use of data may be governed by designated authorities, such as independent research ethics committees or data privacy regulators and the outcomes of approval decisions may be publicly communicated. Data subjects may be given reasonable means to extend or withdraw their consent over time.

Methods of seeking data subject consent are sometimes described as opt-in (affirmative consent) or opt-out (negative consent). Generally, in an opt-out model, if a data subject does not take an action to opt-out, then their data may be approved for research and statistical uses, subject to suitable safeguards. In an opt-in model, a data subject must take an affirmative action to declare that their data may be used for future approved statistical or research uses, subject to suitable safeguards. The way in which the choice is presented to the data subject, whether a default choice is offered, the nature of the response required, and other factors can affect the fair characterisation of a data subject’s response as affirmative or negative.

Other approaches recently proposed in the scientific literature include “adaptive” or “dynamic” models of consent forms, whereby (following the initial “general” consent) participants would be asked to re-consent for any “new” direction of travel/use of their data, potentially using web-based communication tools. This approach is 'dynamic' because it allows interactions over time; it enables participants to consent to new uses of their data or to alter their consent choices in real time as their circumstances change and to have confidence that these changed choices will take effect (Kaye, 2014).

The appropriateness of a more general consent question is under discussion in several countries

The question of the appropriateness of a more general consent question has arisen in several countries (Denmark, Finland, Israel, New Zealand and Sweden). The question to be resolved is whether or not patients could consent today to broadly defined uses of their personal data in the future. A practical example is provided by New Zealand. New Zealand has been collecting blood spots from new-borns since the 1960s. There are a few tests that can be conducted now on the blood; however, it is anticipated that, in the future, there will be additional tests developed that can yield new research possibilities with these samples. This pushes existing legal boundaries about how far does consent extend when the use of data is for a future and not yet specified purpose?

There is a Health Information Governance Framework under development in New Zealand which will provide the health information standards and rules specific to health information. The development of this framework is an initiative of the Health Ministry in collaboration with other stakeholders such as health practitioners, vendors, and consumers. This project includes the development of generic forms to obtain informed consent to the use of personal health data that health care practitioners can give to patients when they first sign up for health services from a primary care doctor. The form will include consent to treatment and consent to the use of their personal health information. The consent will apply to the future use of their personal health information and will not be specific to individual future scientific studies. Given that in New Zealand patients are required to sign up with a primary health care organisation, the initial sign up provides a setting for the administration of such an informed consent form. The consent would be specific to the care provided by the Primary Health Care Organisation (PHC) including medicines and tests. However, should the individual require care provided beyond the PHC organisation, such as hospital or specialist care, a consent form would also be completed by the patient on their first admission/visit to these care providers. If the patient changes PHC organisation, a new consent would be required. As a result, it will be necessary for the ministry to track consent for individual patients over time related to data from different sources.

The question of the appropriateness of a more general consent question has arisen recently in Israel. Israel is developing a national bio bank with cancer cells and which may expand to other samples. The decision taken was that if the consent is written and the patient can clearly understand that they are giving the bio bank authorisation for future research involving their sample under certain conditions, then it would be an acceptable consent. Thus, there would not be study-specific consent. Some of the conditions include that the researcher would only get samples that have been de-identified (where the ID has been encrypted (coded)). This more general consent question that is possible under the *Genetic Information Act* is unusual and exceptional. Under other legislation, consent is specific to the purpose of the data collection.

Patient consent is purpose-specific in Korea. However, the purpose of use is not limited to a specific defined project. As a result, the data may be approved for use in future projects that have not been specified at the time the consent is given. Under law when consent is sought it must include the following: 1) The purpose for which personal information is collected and used; 2) Items of personal information to be collected; 3) Period for which personal information is held and used; 4) Fact that a subject of information has a right to not consent and details of any disadvantage due to his/her rejection to give consent.

In Denmark, there is increasing development of data related to biological materials (genetic data) and use of these data in linkage projects with other health care data. Issues of re-identification risk for biological data are in discussion in Denmark. There is also debate currently regarding the consent for use of biological data and whether such consent should be broad, as it is currently, or become study-specific. Denmark has a long history (30 years) of undertaking population-level data linkages studies. Denmark also has a Health Research Ethics Committee to approve projects involving the linkage of biological data and other registries. Thus it has the legislation and the organisation to support linkages of biological data with health and social data on a population scale. The SSI has a bio bank with data covering 60-70% of the population. Data linkage projects involving these data have received ethical committee approval.

Within Sweden there has been a commission created to examine the issue of whether or not there could be a more prospective consent question in certain cases, such as for biological data submitted to biobanks. Currently, the opinion of the Data Protection Authority is that a non-specific consent to research is not allowed and individuals cannot consent for their data to be used in research projects in the future that cannot be specifically described. The government of Sweden might want to allow certain exceptions to this general prohibition opinion and that is the motivation for creating the commission. In its report of June 2014, the commission proposed new legislation allowing non-specific consent in research databases. The *Patient Data Act* and the regulations of the Data Protection Authority require informed consent of patients for their data to be included within Quality Registers. In this case, patients have an opt-out form of consent. There are no other registries that offer an opt-out consent. There are discussions within Sweden as to whether or not patients are really aware of their option to opt-out of quality registries.

In Finland, the relatively new *Biobank Act* (in force as per 1 September 2013) enables asking for a bio bank-specific consent for inclusion of samples and data for several unspecified future research purposes in a specified bio bank infrastructure. Under the *Biobank Act*, the person shall be adequately informed about the owner of the bio bank and about the bio bank which stores the samples and data, as these may be different bodies. The registered field of activities of the bio bank serves as a limitation on the use of its samples and data. Further, the person to be recruited shall be informed of the general nature of biobanking and potential risks, the purposes of the collection and storage, and the voluntariness of consent and the right to cancel or limit the consent at any time. The person may also give consent that his personal data may be disclosed, or be linked to register data. In 2015, nine publicly funded biobanks will be operational. Most of them are joint ventures of university hospitals and universities and regional health care providers. In addition, THL has established a population-based nationwide bio bank. Under the *Biobank Act*, patients will be able to use the Internet portal (*Act on the Electronic Health Record System*) to provide and later to modify their bio bank consent. The laws on health registers are subject to reforms in the near future to improve their safe use for various purposes.

In the United States, HIPAA applies to information and does not apply to physical samples (blood, tissues). There is a separate patient consent for the storage of samples. A broad consent that enables the tissues to be analysed for future research is permissible in the United States. There has been, however, some controversy over the storage of blood spots for the screening of new-borns. Some states allow it and others do not. There are groups concerned about the storage of DNA and want uses of the blood spot samples restricted.

Public opinion research in the United States indicates that the population sees value from health research but also wants to be consulted about the uses of their personal health data.

In the United Kingdom, an opt-out consent model has been developed regarding the use of personal health data for statistics and research in the public interest. This model enables patients to express their choice regarding the future uses of data about them that is within health administrative and clinical databases. Individuals opt-out by expressing to their general practitioner that they do not wish for data about them that is held in national health care datasets to be de-identified and used in statistical or research projects. This model allows patients to make a choice about future uses of their health care data that can be administered in a practical manner and that doesn't require patients to be consulted and decide upon each of the many individual studies that may be proposed. This model provides for patient choice in a way that is not possible under legislative frameworks that enable statistics and research uses of data by exemption to patient consent requirements or by legal authority.

Data sharing for the purpose of research or statistics

Where the secondary use of personal health data for statistics and research is legally permitted, there can be an accompanying need for data sharing. Data sharing may be needed in order to develop a new dataset or registry on an on-going basis or to conduct an ad-hoc project. The data shared may need to be in an identifiable format in order for datasets to be linked together at the record level.

Fourteen countries have three or more custodians of the key national health datasets investigated in this study (Table 3.4). Further, if a broader spectrum of health care datasets had been studied, the number of national dataset custodians would have been even higher in some countries. Countries with multiple dataset custodians have a greater need for data sharing mechanisms for the development of statistics and research requiring data linkages of national datasets.

Table 3.4. Number of custodians of national datasets

1-2 custodians	Czech Republic, Iceland, Italy, Japan, Switzerland, Turkey, UK England, UK Scotland
3-4 custodians	Canada, Denmark, Finland, Israel, New Zealand, Singapore, Spain, Sweden, United States
5-6 custodians	Korea, UK Wales
7 or more custodians	Ireland, Netherlands, Norway

Source: Authors own calculations based on the results of this study.

Countries differ with respect to whether or not it is ever legally permissible for a data custodian to share identifiable personal health data for statistical or research purposes. Further, in some countries the rules regarding permission for data sharing differ depending on the sector of employment of the organisation or researcher with whom the data would be shared.

Countries where identifiable personal health data may be shared with conditions

National experts in Denmark, New Zealand, Norway, Finland, Sweden, the United Kingdom and the United States stated that the sharing of identifiable data may be permitted for research or statistical purposes subject to approval processes and data security controls. In these countries, the same rules apply to applicants from the government sector, the academic or non-profit sector and the commercial sector, as long as the purpose of the data use is research.

An example from Finland illustrates a situation where identifiable data are required by external researchers. In this example, there is to be a data linkage performed between the THL registry data and the electronic clinical records held at the local or regional levels.

In Sweden, risks to data subjects of any sharing of identifiable data must be minimised and risk factors are evaluated, such as whether the data receiver is under the same legal requirements for privacy protection as the data custodian and whether there is any risk that data could be handed over to a third party. For commercial entities, approval is more likely to be granted when the commercial entity has collaborated with a university researcher and it is the university researcher who would be approved access to the identifiable data. The university researcher would be legally bound to only provide research results (non-identifiable) back to the commercial entity.

In Denmark, a national applicant can be approved access to identifiable health data if their project justifies access to this level of data and the project is approved. In Denmark, the Statens Serum Institut (SSI) shares with regions and municipalities, their own data in an identifiable format without a requirement for any external approvals. If a region or municipality wanted to undertake a national project with personal health data, however, then they would follow the same approval process as would any other third parties.

In the United States, legislation does not restrict applications for a research use of identifiable data by any sector of society. However, there is a restriction on commercial applications to use data for marketing purposes. Marketing is not an exemption under HIPAA and therefore marketing uses require patient consent or patient authorisation. Research ethics boards work hard to distinguish among applications for data access to protect themselves from a violation of the law. Key national health datasets in the United States, however, are often governed under legislations specific to public authorities. The National Center for Health Statistics (NCHS) is governed by its own legislation. For record linkages between NCHS datasets and datasets of other agencies, agreements are formed between agencies of mutual benefit. It does not share identifiable microdata with academics or other third parties wanting access to data. Instead there are mechanisms enabling requests for data linkages to mortality data and for access to de-identified microdata. Similarly, the Agency for Healthcare Research and Quality (AHRQ) does not share identifiable data with third parties, and offers data linkage services and access to de-identified data.

Countries where identifiable data are shared with organisations legally authorised to receive it

In several countries, national data custodians indicated that most or all identifiable health data may only be shared with patient consent or in cases where there is a specific legal authorisation for the sharing.

If legally authorised, data sharing among government entities can occur in Israel for the purposes of undertaking a data linkage. Where possible, however, this is avoided by de-identifying the microdata and including an encrypted ID number that can be used to link the data. In exceptional cases, such as when patients would be contacted as part of the study, identifiable data may be shared among ministries. There is also an exception, under law, for researchers connected to a university hospital/medical faculty. Other researchers cannot be approved access to identifiable data, with the possible exception that access may be granted to identifiable birth or death data.

Under law in Korea, identifiable personal health data may be shared among public authorities. There is no sharing of identifiable personal health data from HIRA with the academic/non-profit sector or the commercial sector. However, in Korea, different sources of data are collected under the authority of different legislations. For example there is a *Cancer Management Act,* a *Genetic Management Act* and an *Organ Transplant Act*. Each legislation has clauses related to data sharing. Many data holdings are supervised under the *Statistics Act*. The *Statistics Act* permits the sharing of identifiable data for Research and Development purposes that is inclusive of research conducted for academic purposes. Thus, some national data holdings in Korea may be shared in identifiable format with researchers from the academic/non-profit sector.

In Iceland, under law, identifiable data are never shared with external organisations unless they have a legal authorisation to receive it. For example, under law, the Icelandic Health Insurance or Social Insurance Administration can receive identifiable medical records. Also, under law, physicians with a smart card can access a medications database that enables them to check their patient's current medications to prevent potential adverse drug reactions or to check for prescriptions for addictive substances.

In Canada, the sharing of identifiable personal health data are subject to any applicable national, provincial or territorial legislative requirements under which the data were collected originally. At CIHI, identifiable personal health data can be shared without consent only if the disclosure is legally authorised. In practice, there has been no sharing with the academic/non-profit or commercial sectors. Cleaned and validated data are, however, often returned to their original "data owners" for their own use.

Countries with legislation that prohibits data sharing or data linkages for research and statistics

A number of countries have legislation that prohibits the sharing of identifiable data for the purpose of research and statistics (Switzerland, Netherlands, Czech Republic and Spain). This does not mean, however, that the linkage of data for approved projects is impossible in all of these countries. In some countries, mechanisms have been found for public entity co-operation that permits high quality linkages for approved purposes.

By law, the Federal Statistical Office (FSO) in Switzerland does not share identifiable data with any other organisation and it is the only organisation that can link data in its custody. Researchers can request a linkage of the data holdings of the FSO and can request a linkage of FSO datasets to a dataset collected by another party. In this case, the approved data requestor would submit their data to the FSO and the FSO would conduct the linkage and de-identify the data.

Similarly, in the Netherlands, under the *Statistics Act*, Statistics Netherlands cannot share identifiable microdata with third parties. Health care providers are also not legally authorised to share identifiable health data for projects. Data linkage projects can still take place in the Netherlands, however, as the law allows a trusted third party to conduct them. Data linkage projects involving Statistics Netherlands data are conducted by them. Other data linkage projects are conducted by a trusted third party, Health Care TTP that links data and provides access to de-identified data.

In the Czech Republic, identifiable personal health data are not shared among public authorities, with academic or non-profit sector researchers nor with commercial organisations. The only lawful sharing of identifiable data is as part of health documentation between providers of health care for an individual patient. It is still possible, however, for data linkages to occur involving the data holdings of more than one organisation. As an example, there is an initiative to link data held by insurance companies (medical claims) to data held by the IHIS. To undertake this work, the civil registry ID numbers are first encrypted following the same encryption algorithm (Hash Function) and then the encrypted ID numbers are used to link the data. It is possible to provide approved researchers (academic) with microdata that has been anonymised by removing direct identifiers and providing only the encrypted ID numbers.

In Spain, the sharing of identifiable personal health data among public authorities is an integral part of the information system for patient care and, by law; patient consent is not required for this sharing to occur. However, patient consent is required before national health care datasets could be linked for a statistical or research purpose and obtaining such consent has not been feasible. In Spain, identifiable personal health data is not shared with members of the academic or non-profit sector, or with members of the commercial sector.

Sharing and access to de-identified data

The sharing of de-identified microdata for approved research and statistics uses is presumed to carry less risk to data subject’s privacy. Such de-identification can be very thorough resulting in a determination that the microdata are anonymised; but the degree of de-identification and the resulting re-identification risk varies among countries and among data sources within countries (see Chapter 7). In general, access to de-identified data that still caries re-identification risk is subject to similar approval processes and data security controls as identifiable data.

Countries where access to de-identified data may be approved for applicants from throughout the society

National custodians in several countries will review applications for access to de-identified microdata from applicants from all sectors of society (Denmark, Finland, Korea, Switzerland and United Kingdom). Applications are reviewed on their merits including that the uses of data are acceptable uses, such as research and statistical purposes, and the conditions for data privacy, confidentiality and security are all met.

In Denmark, Finland, Korea and United Kingdom, researchers from all sectors of the economy can apply for access to de-identified microdata. The approval process depends on the project being a statistical or research use of data, not the applicants’ sector of employment.

In Korea, de-identified data may be lawfully shared with researchers working in the public and in the academic and non-profit sectors. Recently, there has been an opportunity introduced for researchers in the commercial sector to also be approved access to de-identified data. As of November 2013, a new law enables private sector organisations in Korea the possibility to be approved access to public data holdings. The rules for approval for access to de-identified microdata will be the same for private sector researchers as they are now for public sector/academic researchers. In Ontario Canada, a pilot study is underway at the Institute for Clinical Evaluative Sciences to evaluate the provision of access to de-identified microdata to applicants from the commercial sector for research or statistical uses of data through a secure remote data access facility (see Chapter 8).

In Switzerland, all organisations, whether public, academic or for-profit follow the same application and approval process for access to de-identified microdata from the Federal Statistical Office. Applicants for access to data must clearly explain the research project to be conducted. If the request is to conduct research, then they can be approved. For example, there are researchers working as private consultants. The FSO's view is that there is no justification for setting exclusion based on organisational status but judges all applicants on the merits of their project. The FSO is, however, more cautious regarding projects to be undertaken by Interest Groups. For example, when a hospital requests data for all hospitals and could use the data to gain a competitive advantage. In such cases, the FSO would take further steps to reduce the re-identification risk of the data and may not approve access to microdata.

Countries restricting access to de-identified data from commercial sector applicants

Some countries make a distinction between applicants from the government and university/non-profit sectors and applicants from the commercial sector [Canada, Iceland, United Kingdom (Wales), Netherlands, United States]. In these countries, only the consent of the data subjects permits approval to access de-identified microdata for applicants from the commercial sector.

In Iceland, members of the academic/non-profit and government sectors may be approved access to de-identified registries and de-identified linked data. Applicants from the commercial sector cannot be approved and may only receive tabulated (aggregated) data. However, if the commercial-sector applicant has the informed consent of the data subject, then access to de-identified registry or linked data may be approved. An example is the DeCode project within which a pharmaceutical company has obtained the informed consent of data subjects and is able to be approved access to de-identified linked data for statistical analysis.

In the United Kingdom (Wales), the SAIL project provides approved academic and non-profit sector researchers with opportunities to access de-identified microdata including linked microdata via the SAIL remote data access system. There is a component connected to the UK SAIL project that assists commercial entities in accessing the de-identified SAIL data. It is called the E-Health Industries Innovation Centre. Through this initiative, SAIL provides software and app developers with access to synthetic data so that they may develop and test products or training programmes related to SAIL data, without risk to individual’s data privacy.

The academic and non-profit sector in Canada may apply for and be approved access to de-identified personal health data from the Canadian Institute for Health Information (CIHI). Also, if a data linkage has been approved, it is possible for an academic/non-profit researcher to submit their dataset to CIHI and then to have CIHI link their dataset to a CIHI dataset. CIHI would then share a de-identified linked dataset with the researcher. Disclosures to third-party data requesters must be consistent with CIHI’s mandate and core functions and facilitate health or health services research and/or analysis. The researcher must apply for access and must present a justification for each variable that the researcher needs.

Examples where de-identified microdata are not shared but mechanisms are available for data access for research and statistics

Some countries provided examples where there was no sharing of de-identified microdata with external applicants permitted unless the sharing is authorised by law or by the consent of the patients, but, mechanisms were found for the provision of access to the data (Netherlands, United States, Canada, Japan and Singapore). Mechanisms permitting access to data without sharing data are explored in greater detail in Chapter 8.

In the Netherlands, de-identified microdata are not shared but access is provided for approved users through a secure remote data access facility. Public authorities can be granted an on-going right to access certain data via the secure facility and not request approval on a project-by-project basis. Researchers from the academic, non-profit and commercial sectors can also be approved access to data via the secure facility on a project-by-project basis.

In the United States, the NCHS offers data linkage services for linkages to mortality data and mechanisms for access to de-identified microdata to public authorities, academics/non-profit researchers and for research projects undertaken by commercial entities. In all cases, the purpose of the proposed project must be statistical to be approved. Market research is not permitted. For a data linkage, the researcher must submit their cohort of data to the NCHS who conducts the linkage on their behalf, then removes direct identifiers from the data and enables the data to be analysed by the researcher in either a secure supervised research data centre or via a secure remote data access system. Applicants may also request a geographic linkage where contextual data are merged with individual person records based on residential geography. Data that have been merged in this manner are accessed within the research data centre.

Similarly, the United States AHRQ will not provide de-identified linked microdata to an applicant. Instead applicants can be permitted to analyse the data within the secure AHRQ facility. Likewise, Statistics Canada provides access to de-identified data carrying a re-identification risk only via its secure research data centres or its remote data access service.

Only recently has access to de-identified national microdata from the Insurance Bureau in Japan been provided to researchers. Public servants, university-based and research institute researchers, and researchers supported by public research funds can request access to these data.

In Singapore, the Ministry of Health will provide access to linked and de-identified data to approved researchers within its microdata access lab. The ministry will not give identifiable data to external researchers that would add to information they have about individual patients and therefore they must conduct their research within the lab. Even though the data within the lab has been anonymised, it would not be difficult for an external health care institution to re-identify it if it were provided to them. The lab is a neutral zone for safe data access.

Examples where barriers to sharing and access have not been addressed

In the Czech Republic, national privacy law requires that data shared for research purposes is anonymised. Thus there is room for the sharing of de-identified microdata for approved research projects with public authorities and with academic or non-profit sector researchers. However, there are no perceived incentives for national data custodians to share data with research organisations and there is no comprehensive policy encouraging sharing. Further, national custodians that approve the sharing of data for research face possible sanctions from the Office for Personal Information Protection if there is a complaint and if the Office determines that the risk mitigations put in place to protect the data were insufficient. Researchers often face “soft” barriers, such as fees to prepare the data that are unaffordable. Only aggregated data are shared with the commercial sector.

Data linkages with national health care data in Spain are not permitted for research purposes by any sector of society. Researchers from all sectors of the economy can apply for access to existing datasets of de-identified microdata.

In Italy, for data from the Ministry of Health, it is only possible to request custom aggregate data tables. When personal health data are to be transferred from the original data collector, they must first be anonymised and then they must only be used to disseminate aggregated statistics. There is no researcher access to anonymised record-level data. A recent law introduced the possibility to link different sources of data by using a consistent algorithm to generate an anonymised person ID number which is then used to link the data. The Ministry of Health is responsible for the data linkage. The National Agency for Regional Health Services (AGENAS) may use the de-identified linked data to evaluate the outcomes of the national health system. Record-level data are available describing health expenditures. In this case the data are provided by service provider, such as data disseminated by hospital and by expenditure on medicines and products, for example, expenditures on medical devices.

Foreign applicants for access to data

Some countries make no distinction between foreign and domestic applicants for secondary data use, subjecting both to the same set of rules. Nonetheless, many countries are reticent to approve foreign applications for access to data, due to the inability to impose sanctions on a foreign entity for non-compliance with legal requirements or with the requirements within their data sharing agreement. Some countries will not consider any foreign applications; some will consider only applications for access to de-identified personal health data; while others will consider the approval of the sharing of identifiable personal health data if there is a strong justification for the project.

Countries that may permit the sharing of health microdata with a foreign entity

In Europe, the *European Directive 95/46* applies to countries of the European Economic Area (EEA), which includes all EU countries and Iceland, Liechtenstein and Norway. The directive enables the free movement of personal data in Europe and states that personal data can only be transferred to countries outside the European Union and the EEA when an adequate level of protection is guaranteed. With the EEA, all countries would have the same protection of privacy as was required by the directive. As a result, the European countries participating in this study have a clear and similar interpretation of data sharing requirements with foreign entities. Data may be shared if they are fully anonymised, such as aggregated data. If data are identifiable or de-identified but still carry a re-identification risk, then the data privacy protection legislation in the applicant’s country must be evaluated as providing adequate protection.

In the United Kingdom, for example, restrictions on data sharing with foreign entities are described in the *Data Protection Act*. This act requires that processing is only undertaken outside the European Economic Area (EEA) if there are guarantees of a satisfactory level of protection for personal data. Thus it is possible under law for approved sharing of identifiable personal health data. De-identified data can also be shared but a distinction is drawn between completely anonymised information and microdata that has had direct identifiers supressed but still carries a re-identification risk. Such data still require that the foreign country guarantees a satisfactory level of protection for personal health data.

Sweden provided a further caveat. In Sweden, while a foreign public authority may be approved access to de-identified microdata if they are under similar legislative protections to an EU or EEA country, further criteria for approval would be the interest of the Swedish state in the project proposed. It is preferred if the foreign authority can collaborate with a Swedish researcher so that access to microdata can take place within Sweden and only aggregated and non-confidential study results are shared with the foreign authority.

Multi-country projects that require data sharing are still rare in Switzerland and it can be difficult to judge if the legal framework of a non-European state is equivalent or not. Most multi-country projects are parallel studies where Swiss researchers analyse the Swiss data and report only non-confidential statistical results or aggregated data outside of the country.

Iceland indicated that the Data Protection Authority maintains a list of countries where data sharing is permissible and this list includes all countries following the *European Directive 95/46*. While most approved sharing involves anonymised or de-identified data, Iceland provided an example where the sharing of identifiable data was required. Nordic countries have a legal obligation to inform one another of the identity of health care professionals in receipt of a formal reprimand. The purpose of this requirement is to inform other health systems that may consider the individual for employment.

In Denmark, the sharing of de-identified data with a foreign applicant requires approval of the Data Protection Authority and can involve a signed agreement among the countries involved. For example, there is a signed agreement to a data sharing arrangement between Denmark and the European Union, Switzerland and Norway.

Evaluating the adequacy of foreign laws

European countries shared examples where project approval decisions have been complicated by a lack of information regarding whether the legislations protecting personal health data in the foreign country of the applicant provide an adequate level of protection when compared with the national laws; and where the legislative protections of the country of the applicant have been found to be inadequate.

Finland shared an example where a researcher from Australia requested access to de-identified microdata for a project. The researcher and THL worked together to explain to the Data Protection Authority why the detailed data needed to be provided to the Australian Researcher. The document went back and forth to the DPA several times before the DPA could be satisfied to release the de-identified data to the researcher. This process took 3-4 months.

In a second example from Finland, a researcher from the United States was seeking access to de-identified microdata for a project. In this case, the legal framework in the United States was found to be very different from that of Europe. If a US institution is included within a safe harbour agreement, which means that it has been verified to have similar data protections to Europe, then the institution can be treated similarly to a European applicant. However, the experience of THL was that many institutions doing credible scientific research in the United States were not included in the safe harbour agreement. As a result, THL decided not to share data with US institutions outside of the safe harbour agreement. US researchers have been granted access to de-identified data only when they are able to work in Finland and access the data on-site at THL.

In an example from the United Kingdom (Wales) SAIL project, it was possible to grant an applicant from Australia access to de-identified microdata because there was no need to transfer the data to them. Instead, the approved researcher is able to work with the data within a secure remote data access system, just as would any domestic applicant for access to de-identified data.

Treatment of foreign applicants by Non-European countries

Similar to Europe, Israel will consider foreign applicants from countries within the European Union or whose data protection legislations are similar to those of the European Union. New Zealand will also consider foreign applications for access to data where the country's privacy legislation offers equivalent protections to that of New Zealand.

Further, New Zealand shared two examples where it was necessary to arrange for the sharing of identifiable personal health data across borders. First, there was a need for New Zealand data holders to be able to access cloud computing services offered by service providers in Australia and vice versa. To enable the sharing of cloud computing service providers for the processing of identifiable personal health data, New Zealand and Australia developed cloud computing guidelines which impose the same requirements for data security and protection on organisations in both countries. Second, there has been the need to share identifiable data for cancer research, as there is high population mobility between Australia and New Zealand, as well as cross-border care seeking. For such research to be approved there must be significant benefits of the research results for New Zealand and the requesting researcher must have the informed consent of the data subjects.

In the United States, there is no distinction under HIPAA for foreign entities requesting access to data. Foreign researchers can apply for and receive access to identifiable microdata. Such a disclosure requires the approval of a research ethics board as it would for any domestic applicant. Disclosures may, however, be prohibited by policy.

In the United States, in the past, a foreigner could apply for access to de-identified microdata within the NCHS Andre secure remote data access system. However, this practice ended when the *Confidential Information Protection and Statistical Efficiency Act* (CIPSEA) entered into force in 2011. This law applies to all statistical agencies and statistical units at the federal level and it requires them to supervise and control the use of the data they hold. The interpretation of the law was that access to data by foreigners via Andre might not constitute sufficient supervision and control. Foreign applicants remain welcome to follow the same approval process as domestic applicants, but they can only be granted access to data within the Research Data Centres. Similarly, the AHRQ also offers foreign applicants access to de-identified microdata within its facility only.

In Canada, disclosure of de-identified health data are subject to any applicable jurisdictional legislative requirements under which the data were collected originally. CIHI may disclose de-identified data to recipients located outside of Canada except where prohibited by law or by agreement. All disclosures must be reviewed internally by CIHI and approved by CIHI's President and CEO. In some cases, approval from the appropriate Ministry of Health may also be required. Given the additional risk associated with providing data outside the country, it may be necessary to provide further data treatment to reduce re-identification risk, such as less geographic level. The data disclosure agreement and associated data security obligations would be the same as for a domestic applicant.

The principle in Korea is to be restrictive on the approval of access to de-identified data from foreign applicants. Data related to the medical services received in Korea is viewed as too sensitive to be shared outside of the country. However, it may be possible to approve the sharing of a sample of the population. In general, the data would only be shared with a foreign government or international organisation when required by treaty or another international agreement.

Legislation in Singapore protects patients in Singapore. If data subjects have provided consent, then it is clear that data sharing with a foreign entity could be approved. The concern is how a data breach in a foreign country would be addressed. In cases where there is not consent of data subjects, it may be possible to share anonymised data, but the concern is how the terms of the data sharing agreement with a foreign entity could be enforced. This is not a clearly defined area and decisions on project approval involving foreign entities is determined on a case by case basis and depends on the risk of re-identification and the protections of the security of the data that would be in place.

Data sharing challenges among national health dataset custodians

Countries provided a number of examples of obstacles to data sharing among the national authorities in the custody of key datasets that are having a negative impact on the development of statistics and the conduct of research across the pathway of health care. The challenges faced involve differences in legal requirements and data sharing policies among national dataset custodians.

Statistical and other authorities cannot share identifiable data with health ministries – problems and solutions

In Switzerland, as is the case in many countries, the law enabling the national statistical organisation prohibits the sharing of identifiable data with other national organisations or other entities. Data linkages involving the datasets of the Federal Statistical Office must be undertaken by the FSO. Finland and Denmark also signalled that their national statistical authorities cannot share the wealth of socio-demographic information they hold with health ministries for approved statistical or research uses. Instead, data must flow to national statistical authorities to conduct linkages for approved projects.

Several challenges result from this situation. In Spain, even mortality data are not accessible by the Health Ministry for basic monitoring of deaths following treatments and care. This is because the confidentiality of deceased persons is protected. In Finland, there are a few key health care datasets held under the *Statistics Act*, such as home care data and, as a result, the data cannot contribute to registries developed within THL involving data linkage. Statistical Authorities may place further restrictions on access to data they have linked to Health Ministry data. An example from Finland is perturbing the data in the de-identification process to the extent that its utility has been compromised.

A few countries have developed solutions to improve the ability to share data between Statistical and Health authorities. The Directorate of Health in Iceland entered into an agreement with Statistics Iceland wherein each organisation agreed to use a common algorithm to encrypt the patient identifying number so that data could be shared for a linkage project based on the encrypted ID. Staff members from both organisations would supervise the linkage and staff members would personally transport the encrypted data to the linking organisation. The agreement would enable either the Directorate of Health or Statistics Iceland to be the organisation conducting the linkage, however, the agreement has yet to be tested in practice. In the United Kingdom, England reported that viable data sharing mechanisms have been found between statistical and health authorities. In England, the national statistical authority is legally authorised to share data but only if the data will be used for a statistical purpose.

New Zealand also signalled challenges negotiating data sharing arrangements among public authorities and signalled that the drafting of data sharing agreements with other national authorities is a mechanism to overcome barriers to data sharing among public authorities and data linkage projects involving data from multiple organisations.

Negotiating data sharing agreements among public authorities can be slow or impossible

Several countries reported difficulties negotiating data sharing arrangements among custodians of key data. Turkey described that legal regulations and a lack of interagency co-operation limits data sharing among public authorities. Singapore described challenges to negotiating data sharing arrangements with public agencies because these agencies are separate legal entities (corporations) and are not part of the government. Singapore also noted that the lack of a trusted third party to carry out data linkages and the lack of a framework for data anonymisation makes it difficult to overcome this barrier. However, efforts are underway in Singapore to develop a framework to facilitate and encourage data sharing across public agencies.

In Japan, the different authorising legislations applying at the national level, at regional levels and for public corporations create barriers to data sharing arrangements among national authorities and limit data linkages.

In Norway, current laws do not permit the Ministry of Health to share data with any other legal entity (organisation); however, there may be changes in the future as the law is under review.

Canada noted that Canadian provinces each negotiate a data sharing agreement with the Canadian Institute for Health Information and with Statistics Canada and it is a time-consuming process to negotiate these agreements. It is also time consuming to seek approval for data linkage studies when the data from more than one organisation are involved.

In the United States, the national data sharing arrangements involving NCHS datasets took years to negotiate and involved the legal staff of the participating organisations. Further, as these are on-going data linkage initiatives, each time there is a new wave of NCHS data to be linked, there is also a need to revisit the negotiated agreement with the participating organisations. Over time there are changes to the laws authorising the participating organisations which require changes to the process or the data sharing arrangement. There are also changes in the key personnel responsible for negotiating data sharing arrangements and with changes in personnel come different interpretations of existing laws. Lastly, changes in technology can require changes to the technical requirements that then require redrafting of the agreement. Thus each negotiation requires significant time.

Data sharing challenges involving health care providers' clinical data and national authorities

Several countries signalled a lack of a legal or regulatory obligation for health care providers, such as physicians and hospitals, to contribute to the development of statistics and research to monitor and improve health and health care pathways. A particular challenge involves the extraction and sharing of data from clinical patient records for statistical purposes. Some countries report legal and policy constraints to extracting and sharing data from electronic clinical record systems for national datasets or projects. In other countries there is no distinction made in law regarding the source of personal health data and data may be drawn from electronic patient record systems for statistical and research purposes, subject to the same rules as those applying to any other sources of personal health data, such as administrative records.

Electronic clinical data treated differently under law than other personal health data

In Israel, the *Protection of Privacy Act* has specific sections speaking to the protection of electronic data. Electronic medical records are treated differently than "on paper" medical records – but the same as other electronic sensitive personal information. The differential treatment is due to the greater accessibility of electronic records compared with paper records and hence, a higher privacy risk.

In the Czech Republic, there are discussions about introducing a shared electronic clinical record and there is a project to develop a shared record between 2014 and 2020. Under current law, it is not possible for persons outside of a health care institution to request a data extraction from clinical records. There are indicators developed from data extracted from clinical records within institutions but such indicators are calculated by the data owners only.

In Finland, the *Act on the Electronic Health Record System* applies to all patients and enables creation of a national archive of electronic patient records for all patients. The law requires patients to be able to view their personal health data via the Internet. The Act includes e-prescription records. The law has a complex relationship with other laws regarding personal health data. First, the Act does not permit the electronic health records within the archive to be analysed for research or to produce statistics. However, the *Act on National Registries* permits all data to be used as is needed to create the registries. As a result, data can be extracted from the national archive for the purpose of populating authorised national registries. If there was a public benefit to creating a new registry based on data within the archive, then a new authorising legislation would need to be enacted to enable this registry to be created. The practical issues regarding the requirements of the act on the HER system will become clearer during the next few years as the archive begins and is used.

In Singapore, there are legislations requiring health care institutions to report diseases including The *National Registry of Diseases Act* that requires reporting cases of cancer, AMI, renal failure, stroke; and the *Infectious Diseases Act* that requires reporting of these diseases. When required by these laws, data can be extracted from electronic clinical records. For all other diseases or treatments, the extraction of data from electronic clinical records is legally prohibited. If another disease is to be reported, then the regulation under the *National Registry of Diseases Act* must be modified. The *National Registry of Diseases Act* sets out the framework within which the Ministry of Health can develop regulations specifying reportable diseases. Therefore, adding to the reportable diseases requires changing the regulation and does not require returning to parliament to amend the Act.

In Korea, electronic clinical records were developed and are used by private hospitals. However, public authorities are not accessing or extracting data from these records for monitoring or research because of the provisions of the national law for the protection of information privacy. National level use of data from electronic clinical records is not a current priority.

Health care providers not required or not willing to share data

Iceland also has a national electronic clinical record system and is able to extract data from this system to create national datasets. In Iceland, the Data Protection Authority has documented that there are physicians that are not submitting data when they are requested to do so by national authorities.

In Canada, electronic medical records for primary care encounters are in the custody of the health care providers and the providers have no legal obligation to share data with a national authority. This is why the coverage of primary care data in Canada remains low, as each provider is participating voluntarily in the dataset.

In the Netherlands, there are many different data owners which complicate data sharing arrangements. Data owners have commissions within them that must approve data sharing with a third party. Statistics Netherlands is working actively to improve national health care data. Under their authorising legislation, as soon as government invests public funds in a dataset, Statistics Netherlands has the legal right to access the dataset. Nonetheless there are many health care datasets that do not receive government support and data owners are concerned about their data being used for performance reporting. Primary health care data and hospital data in the Netherlands, for example, are not financed by the government and, as a result, data owners only grant approval for data to be accessed on a project-by-project basis.

In the United Kingdom, England, there are data available on audits of clinical care for surgical specialties that is collected at the level of clinicians. There is sometimes a reluctance to make this audit data available for wider analysis. Concerns include that the data will be incorrectly interpreted. The solution has been to invest in improving the metadata so that the data limitations can be understood. Sometimes underlying concerns about making data available, is the desire to keep the data within the small team that collected it, so that they may have the first opportunity to publish research findings from its analysis. NHS England now funds national clinical audits which are commissioned through the Health Quality Improvement Partnership and many are led by the relevant specialty's Royal College. There are a number of developments to make more clinical audit data public, to extend the range of care covered by clinical audit and to improve the frequency and timeliness of clinical audit data. Evidence from cardiac surgery, where clinical audit data have been published for a number of years is that publication drives up data quality and also leads to significant improvements in the quality of care as it facilitates comparison with peers and sharing of best practices.

New initiative supporting multi-country data sharing – the Farr Institute, United Kingdom

In the United Kingdom, the Farr Institute for Health Informatics Research is a collaboration involving four academic research centres in Swansea, Wales; Edinburgh, Scotland; and Manchester and London, England. The purpose is to increase interest in data linkage based research by improving secure data sharing among the four research centres. All four of these centres have established data processing and management systems and the purpose of Farr is to find ways to facilitate collaborating together for cross-national research within the United Kingdom, and to improve the quality of statistics and research. For example, there is known mobility of patients among countries in the United Kingdom and, without mechanisms to share data, the patient care pathways are fragmented. Farr will be examining differences across UK research centres in project approval processes and requirements and in cost-recovery models to promote best practices and reduce differences to make it easier to undertake cross-national research within the United Kingdom and between the United Kingdom and other countries (see Box 3.4).

Box 3.4. Enabling multi-country projects: The Farr Institute in the United Kingdom

Launched in May 2013, Farr is a virtual institute of health informatics research with four main hubs and 19 different organisations. Farr aims to promote improvements in health information and research through improvements in governance and privacy-protective access to data that strengthen within and between country research. Objectives include delivering high-quality, cutting-edge research linking electronic health data with other forms of research and routinely collected data, as well as building capacity in health informatics research. The Farr Institute aims to provide the physical and electronic infrastructure to facilitate collaboration, support the safe use of patient and research data for medical research, and enable partnerships by providing a physical structure to co-locate NHS organisations, industry, and other UK academic centres.

The project was originally funded through an open call for proposal sponsored by ten different funding organisations including the Medical Research Council and disease-specific foundations. The funding provided was GBP 17.5 million. Subsequent to this funding, which established the network, there was additional funding provided through a separate grant for infrastructure to support the network. The infrastructure funding was provided by the Medical Research Council and was valued at GBP 20 million and was time-limited (with the funds to be spent within one year). The Farr Institute comprises four nodes distributed across the United Kingdom and led from the University of Dundee (Farr Institute @ Scotland), University College London (Farr Institute @ London), University of Manchester (Farr Institute @ HeRC N8) and Swansea University (Farr Institute @ CIPHER).

The data governance rules in Scotland differ from those of England. Farr aims to bridge gaps in the rules to ease the approval of UK-wide projects. At present, when there is a proposal to undertake a project with data from England, Scotland and Wales, a new agreement for data sharing is negotiated. There are numerous approval bodies within each country and within regions within each country, each with their own applications for data access and approval processes. It is complex. Farr aims to both map out the current landscape for access to de-identified data and work with regions and national entities to streamline processes, so that more projects can be done efficiently with appropriate governance approvals. At present, UK-wide projects take one-to-two years to satisfy all governance requirements including preparing applications and obtaining approval from numerous organisations; and getting data from different organisations into an analysable format. There can be as many as 200 different approvals to secure in the worst case. One of the main challenges to streamlining governance is to get agreement among data custodians and approval bodies that the approval decision of a different body would be acceptable.

Another issue to resolve involves data sharing. In some cases, clinicians involved in a data collection are not in agreement with the data being shared for research to be undertaken by others. When the data collection has been financed by the NHS, others view the data as a public good. A possible strategy is to allow the clinicians to have private access to the data for a fixed time period, such as two years, and then allow the data to be available to other approved researchers after that. Standardised guidelines that are publicly available regarding the use of clinical/phenotypic data would be helpful.

While Farr focuses on data sharing and use within the United Kingdom, the work done will improve the ability of the United Kingdom to engage in international projects involving linked data and aims to improve the United Kingdom's ability to be a partner in international projects. Each node of the Farr has work streams on data governance, public engagement, capacity building, infrastructure and research.

For more information: Farr Institute of Health Informatics Research, www.farrinstitute.org/, and ISD Scotland, www.isdscotland.org/Products-and-Services/eDRIS/.

Legislative reforms that are needed or are underway

Legal experts interviewed for this study provided insight into legislative reforms that are needed or are underway to support personal health data protection and data use.

On 1 April 2013, the *Health and Social Care Act* came into effect in England and made significant changes to the management and delivery of health care. Under the Act, the Health and Social Care Information Centre (HSCIC) became an independent body from the Department of Health and became responsible for consolidating national health care data and providing access to that data under controlled conditions. The HSCIC had its own internal review processes to advise it on the collection and use of personal health data, but it did not have the benefit of an independent review committee. Prior to the new Act, the National Information Governance Board (NIGB) provided advice to the Secretary of State and from there to the NHS on information governance including respecting the common-law duty of care. Under the government's new care.data initiative, the HSCIC provides access to de-identified health care data to requestors from all sectors of the economy. The HSCIC approved the release of de-identified health care data to an insurance association and there were media reports that created a public outcry that sensitive data had been disclosed. The Information Commissioner's Office found that the data release was within the law, but safeguards around decision making about the release of health data needed to be strengthened and needed to be made more transparent to the public. The *Care Act* was introduced in May 2014 and strengthened governance of personal health information in the care.data initiative by recognising the Confidentiality Advisory Group (CAG), which is an ethical review board created by the Health Research Authority (HRA), as a statutory body that has the role of providing independent advice to the NHS regarding care.data and to the HSCIC on the release of data, as well as advice to the HRA. The *Care Act* also made it clear that individuals can object to having their data included in care.data.

In Finland, the *Act on the Electronic Health Record System* allows patients to view their record and to opt-out of sharing portions of their record with health care professionals. Health professionals have expressed concerns as they have no information about what type of content has been removed from the record and therefore they do not know if there is missing information that would be pertinent to the best clinical decisions for the patient's care. The act on the EHR system also does not allow clinical data within the EHR archive to be analysed. As a result, even if patients consent to a linkage of clinical data, the archive cannot be used for this purpose. Instead, the clinical record data will have to be extracted from all of the local data holders instead to enable the linkage. Thus, even after the creation of the national EHR archive, because of the limitations of this legislation regarding data use, it will continue to be necessary to provide identifiable data to local/regional centres for data linkages, even identifiable data involving bio bank specimens. The Ministry of Social Welfare and Health is planning one single legislation on the secondary use of electronic patient and client journals in a government committee in 2015-16. In Israel, there is on-going work to develop rules regarding the processing of electronic clinical records under existing laws. There is also a recently enacted legislation related to the collection and de-identification (including encryption of ID numbers) of health data necessary for the Ministry of Health to have a broader programme of monitoring disease and procedure outcomes.

In Iceland, there was a change in the law governing the prescriptions medications database in 2012 to enable patients to access their prescription medicines data via a secure patient data portal. This portal also provides patients with access to their immunisation records. In the longer-term, the plan is to have one portal where patients can access all of their personal health care data that have been authorised.

In New Zealand, there is work underway to examine whether the national privacy legislation should be strengthened to stipulate legal penalties for breaches of compliance, such as fines, and to allow the Privacy Commissioner's Office to have the authority to audit data holders. Currently, the main remedy for breaches of compliance in New Zealand is the impact on the researcher's reputation, remedial action by their employer and the possibility for complaints to be made to a human rights tribunal for investigation. The privacy framework was developed in the late 1980s and 1990s and it is difficult for the framework to cover all situations arising as a result of increases in computing power, smart phones and devices and the level of access to personal health data. These technological advances hold potential for analysis to improve health outcomes but also challenge the consent model within the legal framework. For example, it is a challenge to seek purpose-specific consent from consumers who are contributing to datasets along with millions of people.

As researchers ask to develop products from data that challenge the privacy framework and require increasingly complex work around solutions to approve, there is some discussion about the need to modify the *Privacy Act* in New Zealand. However, there is also reluctance to revisit it as there is always a small vocal minority opposed to any exemptions to consent requirements. A recent challenge is a request to approve an exemption to consent requirements to link health data with justice and education data. While an exemption can be granted, to what degree do such projects risk damaging the trust of health care practitioners, health care institutions and patients?

The cantons in Switzerland have their own laws and some may allow sharing of identifiable data. There is a new national law entering into force in 2014 that will govern the cancer registries at the canton and national levels. This will standardise practices among cantons. The new law will create a legal basis for cancer registries and will enable cantons to continue to create registries and to share data to complete a national cancer registry. The law will cover data governance including sharing, linkage, anonymisation, security and access to data. The current wording of the draft law would not permit custodians of cancer registry data to share data in identifiable format with third parties. The law creates a template that could be followed to introduce legislation authorising other disease registries in the future.

Singapore is at an early stage or considering mechanisms to improve data sharing for research purposes in order to improve the potential value from medical data to the health care sector. The national electronic health record system creates the potential, but the data was collected primarily for patient care. For research uses the questions are with respect to who could access the data, at what level and under what conditions, such as whether or not it is possible to anonymise the data and what method to use. It is possible that it may be necessary to introduce or amend existing legislation, but operational changes, such as including data sharing agreements and reforming IT workflow would also be needed. What is important is to strike a balance that allows the safe usage of confidential health data. The key issue is to protect patient confidentiality while allowing the benefit of analysis of patient data and, in so doing, make an appropriate trade-off.

In the Czech Republic, the law should clearly state that data linkages can take place if they are in the public interest. Health status and health care quality statistics are public interests. The new *Data Protection Regulation* to be introduced by the European Parliament is expected to make it necessary to reform the national law. IHIS would like the law to allow an ID number to be attached to dataset records and be used for approved data linkage studies involving different sources of data. For example, it is not possible now to conduct data linkage projects involving criminality data and health data or for patients with mental diseases and other health data. Also, it is hard to request linkages between health insurance fund data and other health data. This is because the law governing health insurance data does not contain a clear statement that the data may be used for statistical purposes. This is a problem because the health insurance data are necessary, for example, to understand the economic impact of treatment, the cost of health care and to develop pay for performance plans.

Ireland reports uncertainty regarding the interpretation of current legislation and uncertainty about the evolution of national and European legislation governing the protection of personal health data. Uncertainty about the evolution of the European *Data Protection Regulation* and its impact on future legislative reforms was widely indicated by European experts taking part in the telephone interview with the OECD.

Key elements of legislative frameworks supporting privacy-protective uses of health data

To maximise societal benefits and minimise societal risks requires careful development of strong health information governance. There is no doubt that the legislative framework governing health information and the protection of privacy is the most important part of any governance model. This chapter provides an overview of the features of OECD countries legislative frameworks. It sheds light on the strengths and weaknesses of different approaches and a few important overarching themes emerge. Firstly, whether enabled by a broad and prospective patient consent, exemption to patient consent requirements or legal authorisation, many countries are able to approve privacy-protective uses of personal health data that are in the public interest, including the extraction of data from electronic clinical records and dataset linkages. Second, legislative frameworks need careful development to ensure that protection of health information privacy is consistent and that there are no forms of personal data that fall outside of legal protection. Third, legislative frameworks need to take into consideration the burden imposed on citizens of well-meaning efforts to enable them to exercise their privacy rights and to codify mechanisms that are fair and practicable.

The Advisory Panel of Experts on Health Information Infrastructure identified the following features of legislative frameworks as key factors promoting privacy-protective data use:

2. The processing and the secondary use of data for public health, research and statistical purposes is permitted, subject to safeguards specified in the legislative framework for data protection

The legislative framework should:

a) Reflect the basic principles for privacy protection outlined in the OECD Privacy Framework (OECD, 2013).

b) Cover all data sources and all data custodians and processors.

c) Require a fair and transparent project approval process including an independent, multi-disciplinary project approval body.

d) Permit use of personal health data for public health, research and statistics in the public interest, subject to the approval process.

e) Allow the processing of data, whether by consent, exceptions to consent or specific authorisation, for further approved statistical and research projects. Government statistics, and research, are all activities that should be considered, in principle, as legitimate purposes for the further use of data.

f) In situations where patients have the right to opt-out of the inclusion of their data in datasets used for future approved research and statistics, there are practical means to exercise that right, including where available, technology that simplifies the expression and maintenance of patient choices.

g) Allow personal health datasets to be linked for approved uses (record linkage).

h) Permit the sharing of linkable data among public authorities for approved data linkage projects and government statistics.

i) Permit public authorities and/or trusted third parties to securely store keys to the re-identification of data to enable future approved data linkage projects and government statistics.

j) Allow sharing and access to de-identified person-level health data for research or statistical projects by applicants from all sectors of society, subject to an approval process that includes privacy and security safeguards and prevents re-identification.

k) Allow sharing and access to de-identified person-level health data for research and statistical uses by foreign applicants, where the legislative framework in the foreign country adequately meets the standard for data protection of the home country and subject to an approval process that includes privacy and security safeguards and prevents re-identification.

l) Require public reporting of all applications for approval to process personal health data and the approval decisions.

References

Cavoukian, A. (2012), *Operationalizing Privacy by Design: A Guide to Implementing Strong Privacy Practices*, Information and Privacy Commissioner, Ontario, Canada, December.

CHI – Canada Health Infoway (2012), *Privacy and EHR Information Flows in Canada: Common understandings of the Pan-Canadian Health Information Privacy Group*, Version 2.0, Canada Health Infoway, July.

Di Iorio, C.T., F. Carinci and J. Oderkirk (2013), "Health Research and Systems' Governance are at Risk: Should the Right to Data Protection Override Health?", *Journal of Medical Ethics*, http://jme.bmj.com/content/40/7/488.abstract, accessed 27 July 2015.

European Commission (2012), "Proposal for a Regulation of the European Parliament and of the Council on the Protection of Individual with Regard to the Processing of Personal Data and on the Free Movement of Such Data (General Data Protection Regulation)", COM(2012)0011, http://ec.europa.eu/justice/data-protection/document/review2012/com_2012_11_en.pdf, accessed 20 March 2013.

Gellman, B. (2007), "Consent for Disclosure of Health Records: Lessons from the Past", www.worldprivacyforum.org/wp-content/uploads/2007/04/MaineHealthPrivacy1998_Gellman.pdf, accessed 1 August 2014.

Kaye, J. et al. (2014), "Dynamic Consent: A Patient Interface for Twenty-first Century Research Networks", *European Journal of Human Genetics*, Vol. 1-6.

OECD (2013), *OECD Privacy Framework*, Paris, www.oecd.org/sti/ieconomy/oecd_privacy_framework.pdf.

OECD (2009), *OECD Policies for Information Security and Privacy*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264101630-en.

Chapter 4

Open and transparent health information systems

This chapter reviews progress in the openness and transparency of governments regarding the collection and use of personal health data. It takes a close look and examples from countries and emphasises the importance of not only having strong and effective data governance mechanisms but also ensuring that public communications about data governance are effective and reach their intended audience.

Highlights

While there is no doubt that communicating about uses of and safeguards surrounding personal health data is challenging, transparency is essential to maintaining public trust and confidence in government.

Twelve countries have a policy or programme in place to promote open government health data. Often this is part of a public-sector wide initiative. Countries rarely, however, provide the public with a centralised location where they can inform themselves about all of the national health datasets, and, in particular, the national personal health datasets. Some countries provide the public with information regarding approved studies involving the processing of personal health data including dataset linkages. This information increases public transparency about how personal health datasets are being used, by whom and for what objectives.

Countries supporting researcher access to data tend to be more transparent with the public about data access by providing, usually via a website, information about applying for data access, project approval requirements and legal and practical requirements of approved applicants. Such transparency enables the public to understand and scrutinise data access practices and safeguards and offers fairer access to information to potential data users, whether they are located in the country or abroad.

The United Kingdom shares an important lesson-learned about the necessity of public consultation and effective public communications that must accompany national plans to strengthen health information infrastructure and national efforts to develop data governance that maximises societal benefits and minimises societal risks.

Health information systems are developed for the public's benefit, to ensure health care is accessible, of high quality, and affordable and to generate new scientific discoveries to improve therapies, outcomes and overall population health. However, public awareness of the data inputs to the health information system, the users of the system, the uses of the system and the benefits and risks that are associated with the system is often limited.

Increasingly, OECD countries have initiated projects to increase the openness of public-sector data including national health data. This movement aims to improve the accessibility of health data and information about health data in order to be much more transparent about the data collected by public authorities and to promote the re-use of these data to generate new information value for the public benefit and to stimulate innovation. Some limit this initiative to promoting the availability of aggregated health statistics. Others are focusing on improving individual's access to their own records that are held by public sector entities, including electronic health records. Still others view the move toward openness and transparency as inclusive of improving research and innovation and are including strategies to improve secure access to sensitive personal health data for approved uses.

While there can be excellent health datasets available and mechanisms to provide access to data, the benefits of these resources will not be fully realised unless there is open communication about data availability and data access approval processes and approval criteria. Few OECD countries, however, are open and transparent about the availability, accessibility, security and the benefits and risks of personal health data.

Improving openness and transparency in this sensitive area is challenging. The spectrum of public views will be wide-ranging, from positive to negative, regardless of the data governance structure proposed. Public views are also dynamic and will be shaped by culture, past experience and current events. Public awareness also changes the demand for access to data from public authorities and can result in additional resource expenditures that are beyond current budgets. Nonetheless the input of stakeholders and the general public into the governance of personal health data is essential to develop and maintain a health information system that maximises societal benefits and minimises societal risks.

Elements of openness and transparency discussed in this chapter include whole-of-government open data initiatives; efforts to improve transparency about national holdings of personal health data; transparency about researcher access to data; strategies for communicating with key stakeholders and the public; public opinion regarding research and statistical uses of personal health data; and lessons learned about the necessity of transparent public communication.

Open government health data

Countries participating in this study were asked if their government had a policy or a programme in place to promote open government health data. Overall, such initiatives were signalled within twelve countries: Canada, Finland, Iceland, Italy, Korea, New Zealand, Singapore, Sweden, Switzerland, Turkey, the United States and the United Kingdom (Table 4.1). Reasons for these initiatives include to be more transparent to the public and to make it easier for government data to contribute to policy making, administration, consumer knowledge, business innovation, and so on. Each initiative is described in Table 4.1.

Ireland reported an e-government strategy for 2012 to 2015 to make better and more innovative use of information and communications technologies to improve the public-sector customer experience including data accessibility. The initiative does not, however, include health data. In the Czech Republic, the Ministry of the Interior is leading discussions about open government data. The discussions are related to improving citizen access to their own data that are held by multiple public authorities. In Spain, there is an initiative to provide registry-based data in an interactive web-based tool supporting data queries. The initiative is not part of an open health data strategy.

Table 4.1. Open government health data initiatives

	Open data initiative	Website
Canada	The Government of Canada first launched its Open Government Strategy in March 2011. A variety of aggregated data about health and safety topics are included within a government-wide open data website. The website does not provide information about availability of de-identified person-level health datasets.	www.data.gc.ca
Finland	The Act on the Openness of Government Activities (1999) states that in principle all national, regional and local government data should be open to the public unless there are reasons that the data need to be kept secret. The open data strategy is led by the Ministry of Finance and is mentioned in the government's work programme for 2011-2015. The initiative excludes personal health data because it is sensitive and within the category of data that needs to be kept secret.	Government ministries, including THL, make aggregated data available to the public on their websites.
Iceland	The most recent national strategy (2013 to 2016) involves improving access to de-identified person-level health data via a national data warehouse; as well as secure access for individuals to their own personal data through a government portal. Access to medications data are to be included and access to immunisation data are already available via this portal. These initiatives are led by the Directorate of Health, while the overall open public-sector data initiative is led by the Ministry of the Interior. There is a single government website providing data and information on all aspects of society, including health.	www.island.is
Italy	The Open Government Action Plan introduced in 2012 develops open government data for all public authorities including health. The Agency for Digital Italy develops open data guidelines and tracks progress. A web-site provides links to open data including aggregated health data.	www.dati.gov.it
Korea	The Offering and Use Promotion of Public Database Act of 2013 requires national and local governments and sub-agencies to open access to data to the private sector. This includes de-identified person-level data. The Ministry of Security and Public Administration has published a complementary guideline for public authorities on privacy protection in the sharing of public data.	
New Zealand	A directory of publicly-available, non-personal New Zealand government datasets, including health is provided through a single website that provides links to datasets held on other government websites. It is under the responsibility of the Department of Internal Affairs.	www.data.govt.nz
Singapore	An open data initiative applies to data for the whole of government, including health and health care data. The objective is to make data available for citizens and to create economic value from data use. Aggregated and non-confidential health and health care data are publicly disseminated via a public web portal and including metadata to describe the data.	www.data.gov.sg
Sweden	The government has asked authorities to participate in a technical platform and a website that will publish data and information for secondary uses. The work is mainly for data that is outside of the health sector. The open data web-site is to launch on 1 July 2015.	
Switzerland	The Open Government Data Switzerland Project has launched an open government data pilot portal in 2013 which is to remain available until the end of 2014 before a definitive decision is taken on the launch of open government data. The Federal Office of Public Health is participating in the pilot.	www.opendata.admin.ch/en
United Kingdom	An Open Data White Paper set out UK government policy for the openness of public-sector data (UK, 2012). A single web portal provides access to non-confidential data, including data from the health sector, and is led by the Transparency and Open Data team in the Cabinet Office. Within the UK, England has had a comprehensive approach to strategic planning and implementing open health data (see Box 4.1).	www.data.gov.uk
United States	The Health Data Initiative was launched in 2010 with the aim of increasing the value of health data and fostering innovation while, at the same time, demonstrating that privacy can be protected while data are being used to support healthcare innovations and a culture of data sharing and use (DHHS, 2014). It began with 30 health and human services datasets and has expanded to over 1000 datasets. The initiative includes both aggregated data and public-use micro data and is managed by the Department of Health and Human Services.	www.healthdata.gov , www.healthindicators.gov

Aims of open health data initiatives vary

The most common reported aim of an open health data initiative was to increase transparency in and access to aggregated health statistics as part of a public-sector wide initiative (Canada, Finland, Italy, New Zealand, Singapore and Switzerland). The open health data initiatives in Iceland, Korea, the United Kingdom and the United States have a broader set of aims, including improving the availability and use of de-identified microdata. In Iceland, the initiative is also inclusive of an effort to improve secure access for individuals to their own health records, beginning with immunisation records and expanding to prescription medicines.

The open health data initiative in Italy has a secondary aim to improve data exchange among public authorities. The objective is to improve public administration and reduce the burden on persons and organisations by asking once for information needed by more than one public authority, rather than making multiple requests for the same information. Under this objective, identifiable data can be exchanged, if there is an act that authorises the exchange and provides clarity regarding the scope and rules of the exchange. There also must be security around the exchange of the data to protect it. An example is the exchange of medical certificates regarding illnesses which may be shared from the doctor to other organisations authorised to access the data, such as the Social Security Agency.

The open health data initiative in the United Kingdom is a comprehensive initiative that includes strategic planning about the open public-sector data strategy and the strategy for open health data specifically, and includes guidelines and a rating system for included data, and policies to promote freely available data and to ensure that data are open and transparent wherever possible (see Box 4.1).

Box 4.1. Open government health data: United Kingdom, England

In the United Kingdom, England, there is a cross-government policy to promote transparency and open data across the public sector in support of greater public accountability and service user choice and to support economic growth (United Kingdom, 2012). This work is led from the Cabinet Office and includes an Open Government Partnership. Through the open data initiatives, departments are asked to look carefully at the reasons why data have not been made available more widely, particularly if requests through data.gov indicate it should receive a high priority for becoming available.

The UK National Action Plan describes plans for national information infrastructure including the key datasets from the public sector of greatest value and utility, such as key geo-spatial data (where to find a GP, information on costs for services, etc.). The objective in providing access to government data is to stimulate the development of tools and apps from the raw data. The UK National Archives provides open government licences that set out the terms and conditions that must be respected when using open government data. For example, the source of the data used must be acknowledged. There are strong legislations and information governance to support data privacy protection. The open government strategy intends to enable data to be used freely. There is a push across government to make more of its data available in machine readable/reusable format. There is a rating system for the data that have been provided, ranging from level 1 which are data that are not very useable for secondary analysis, such as PDF documents, up to level 5, which are data provided in a useable, flexible format. Most of the data provided by the HSCIC have been given a rating of 3 which indicates that the data are analysable, in an MS Excel or CSV format.

Some government departments had a policy of charging users for access to data and, as a result, resist making the same data freely available. Further, some will use the data to create apps or tools that are sold. However, the government view is that the taxpayer has already paid for the data to be collected and should not be asked to pay a second time. Some of the elements of the strategy to promote open government data have been financed with existing budgets. Other areas have received new funding. All of the funding has come from public budgets.

More work is needed to measure the cost/benefit impact of the open data strategy in the United Kingdom. A disadvantage of making data open and freely available under UK Open Government Licence is a lack of information about who is using the data and about what is the outcome of the use of the data. There is some on-going consultancy work on how to measure financial impacts or savings from providing open data and to describe the potential uses of data to support the sustainability of the NHS. There are, however, examples of the benefits of more open health data. Following the availability of data on prescription medicines in primary health care by dosage by month, a study found unwarranted variation in prescribing patterns and estimated that if all practices used generics appropriately, it would save the NHS GBP 200 million per year.

Transparency about national health datasets

While many countries reported having a whole of government strategy to improve openness and transparency, countries rarely provide the public with a centralised location where they can inform themselves about all of the national health datasets, and, in particular, the national personal health datasets. Transparency about the existence of personal health datasets would greatly enhance public awareness of health data and its uses and would stimulate interest in data-based research. Public information should include a description of datasets' content, uses, custodians, privacy and confidentiality safeguards, application procedures, approval processes and current projects.

The most commonly reported approach to information sharing about personal health datasets was for each data custodian to provide information about the datasets that they are responsible for. This information is typically shared on the organisation's own website.

Exceptionally, in the Netherlands, RIVM hosts a website that provides information about health care data in the Netherlands that are held by different organisations. It provides metadata for all datasets including the data type, the variables within the dataset, the owner of the dataset and how to request access to the data. The website is an initiative of the Ministry of Health and was launched 8-9 years ago.

In Spain, by law, the registration of every registry with personal data or personal identifiers is compulsory. This registry is maintained by the National Agency for Personal Data Protection. The *Data Protection Act* (article 14) gives the public the right to freely consult the registry to inform themselves about personal datasets, the purpose of the data collection and the identity of the data custodian. Within this registry, every dataset is classified according to three levels of risk: high, medium or low.

In the United Kingdom, the government is identifying key datasets government-wide that should become open. There is one government website providing a single portal for access to government data: data.gov. Through the portal you can either find open data or find lists of data that are not open and information about why the data are not open. The Department of Health in England is gathering lists of previously unpublished health data and making it known that these data exist on the website. Sometimes the data have not been made available before because it was just not thought to do so. When data are listed as existing within data.gov, then people using the website can signal if they believe that the data could be useful if they were to be made available.

In Korea, there is a public web portal where the data holdings of public agencies, including HIRA, are listed. This portal is maintained by the Ministry of Security and Public Administration.

The Information Centre for Registry Research in Finland maintains a website sharing information about administrative registries on all topics, including health (www.rekisteritutkimus.fi). The THL (THL.fi) also has a website where all of the THL registries are documented. What is absent thus far from the Information Centre and the THL documentation is the many health-related registries at local and regional levels where the holders of the registries are public authorities (at the local/regional level). There are also other national institutions with important health data, such as Statistics Finland (deaths, socio-economic data) and the National Insurance Institute which has data on prescription medicines and other insured services.

In Switzerland, the Swiss e-Health Organisation is creating a complete listing of electronic health record systems but this work is not yet completed.

In Sweden, information about national quality registers has been added to the information already shared about the National Board of Health and Welfare’s datasets on its website.

Sharing information with the public about approved studies involving personal health data processing

Some countries provide the public with information regarding approved studies involving the processing of personal health data, typically dataset linkages. Such information increases public transparency about how personal health datasets are being used, by whom and for what objectives. It enables the public to scrutinise data uses and it inspires new ideas for projects involving health data for the public benefit.

In Iceland, the Data Protection Authority publishes all approved projects involving the processing of personal health data on its website. In Israel, all medical research must be made public by registering the studies on the ministry’s website. In Korea it is a legal requirement for public institutions to publish regarding the legal basis, purpose, scope, etc. of uses of personal health data in the institution’s official gazette and website. In Finland, the decision to approve a data linkage project with the registries of THL is made public on the organisation’s website. The study plan is not made public in order to protect the researchers’ interests. In Canada, all approved data linkage studies involving the data holdings of Statistics Canada are summarised on the organisation’s website.

In Sweden, information about project approvals involving data from the NBHW is public but approval decisions are not provided to the public via a list or a website. The information is given only if it is requested.

In the Czech Republic, by law, data controllers must notify the Office of Data Protection before personal health data are processed. The Office of Data Protection keeps this information in a registry that describes the purpose of the processing; the categories of personal data, data subjects and data recipients involved; and the period of data retention. The registry is accessible to the public.

Transparency about researcher access to data

There are differences among OECD countries in the degree to which providing data access services to external researchers is prioritised. In some countries, providing such services is a regular, funded activity; while in others, such services are exceptional and limited. Countries supporting researcher access to data tend to be more transparent with the public about data access, such as offering public information on their website about applying for data access, the requirements for project approval, project approval steps and the legal and practical requirements of approved applicants. Such transparency enables the public to understand and scrutinise data access practices and safeguards and offers fairer access to information to potential data users, whether they are located in the country or abroad.

Finland provides comprehensive information about Finnish registries including the contact details of the persons to ask questions about access to data applications and processes on the THL website (THL, 2014). The website provides the application for data processing/access and it also provides information about researchers’ data security requirements such as how to destroy or archive data lawfully once a project ends. There are links to the data protection ombudsman's office and also to the Information Centre for Register Research. There is a telephone line provided where individuals can speak with a staff member to get answers to any questions they may have. This information for researchers has been made available in English for the benefit of fostering cross-national research. Some of the laws are also available in English, but not all of them. More work remains to enable all of the information needed by researchers to be available in English. When THL receives an application from more than one researcher or PhD student for a project that seems to be similar, THL makes an effort to connect the researchers with one another to see if they may be able to collaborate. Otherwise, one researcher may publish in a high profile journal while the second researcher is then not able to publish because their work is too similar.

In Iceland, guidelines for researchers are available on the Directorate of Health's website. These guidelines are under revision now in order to help to further clarify the requirements of researchers and the process and steps in applying for project approval. The application will also now offer a check list of datasets and variables to make it easier for researchers to convey their data processing needs. Guidelines are also available on the websites of the Data Protection Authority and the Bioethics Committee. The Directorate is considering making their guidelines for researchers available in English. The Directorate also offers guidance to researchers from their staff. Applicants for data linkage services usually consult the Directorate staff before they begin to prepare their applications and they are guided regarding the elements to include in their applications and advised against elements that would never be approved. As a result of this support, most applications are approved by the Bioethics Committee and the Data Protection Authority.

In Sweden, the criteria for project approval are available on the NBHW website. The website provides links to the legislation/rules and also provides a step-by-step description of the process to follow when conducting research with NBHW data. There is information about protecting data secrecy and about the services for researchers that the NBHW provides. There is an application form for data access/processing on the website that researchers use to submit their data request.

In the United States, the NCHS provides a step-by-step guide on its website for researchers regarding applying for access to restricted-use data through its Research Data Center. The website describes NCHS restricted data, modes for data access, the proposal process, data confidentiality policies and requirements (including a confidentiality orientation), and information regarding disclosure review (CDC, 2015). On this website there is also a template for the request for access to data as well as an example of a completed application. The NCHS staffs also works closely with researchers to support them in refining their applications so that they can be approved.

In Denmark, the Data Protection Authority (DPA) provides information to applicants for personal data processing on their website. The SSI provides advice to researchers regarding SSI registry data. The SSI provides an application form on its website to be completed by researchers seeking access to data. The application form asks about the proposed objectives, data and methods of the project and the dataset variables needed. In Denmark, the DPA will allow researchers, for a limited time period, to be allowed access to a more broadly defined set of data in order to explore the data and more precisely identify the data they will request in their final application for access to data through SSI.

In the United Kingdom, the Wales SAIL project provides an application form on their website to be completed by researchers seeking access to data. It asks about the proposed objectives, data and methods of the project and dataset variables needed. Upon application, SAIL verifies the feasibility of the proposed project and, if needed, discusses modifications to the proposal with the researcher. SAIL also supports researchers by informing them when they have proposed a project similar to that of another researcher. The SAIL project also offers training courses for external researchers. There are two five-day courses entitled Introductory and Advanced Analysis of Linked Health Data. Developed by academics from the University of Western Australia, the courses are delivered by Swansea University in Wales. There is also a SAIL user's forum that enables all users of SAIL data to discuss methods and experiences. SAIL is considering developing a web-based researcher training course but is challenged because of the range of knowledge and experience among different research applicants.

In Canada, the CIHI Best Practice' Guidelines for Managing the Disclosure of De-Identified Health Information sets out how requests for data will be handled including approval, de-identification and disclosure and these guidelines are publicly available (CIHI, 2014). Further the CIHI website provides detailed information to guide researchers through the process of requesting access to data and informs them about the relevant fees. CIHI also provides support to applicants to guide them toward preparing an application that could be approved. Similarly, Statistics Canada also makes publicly available its policy directive on data linkage which provides an explanation of the criteria that is followed to approve a data linkage application and provides information on the possibility to request a data linkage for an epidemiological study.

In Korea, the HIRA website has information about requesting access to data and the criteria for approval. There is a template for data access requests provided on the website and researchers can submit their request on-line or by other means. There is also a training programme to help researchers understand the new remote data access system.

In Spain, the criteria for access approval are provided on the MOH website and a template is provided for requesting data. The request can be submitted via the website or by another means.

In Switzerland, the FSO website provides information about the datasets but the website is in need of modernisation. Information about data linkages is very difficult to find. There is an information officer for health that answers e-mail and telephone inquiries if people contact the FSO. There are resource constraints that prevent the website from being revised.

In Israel, researchers can request a data linkage be undertaken by the ministry, with the de-identified data provided back for analysis. However, such linkages would only be approved where the ministry agreed that the project was in its interest, as the approval of the request would consume resources. It is also not easy to find information on the ministry website information about how to request access to data or to request a data linkage, as the promotion of these services would increase the volume of requests and require resources.

In Japan, access to Insurance Bureau data for research is relatively recent and about 10-15 applications for access to data are considered for approval every six months. The Insurance Bureau has two physicians on staff that review applications for data access and work with researchers to refine their applications and prepare them for presentation to the Board for approval. Assistance is needed because researchers are still learning about the insurance data and what it can and cannot be used for. The two staff members are also the only persons involved in responding to requests for information.

In Singapore, the owner of each registry or database each has its own data request form. The form includes the information needed to process the data request and describes the standard terms and conditions for data disclosure. The criteria for project approval are not published. In practice, the data requestor discusses their request with the data custodian regarding the specifics of their project and the data sharing policies and environment specific to their case. The data custodian works with the data requestor to make a decision and to determine the most feasible way to provide the data needed. The requirements for approval differ by database and when linkage is required, the terms of each database involved must be respected. At this point, it isn't possible to create general terms that would apply to all data requests. However, the website for the Ministry of Health makes clear that data requestors should approach the ministry to discuss their data request.

Public opinion about data uses

It is difficult to assess the degree to which populations are enthusiastic about or have concerns about statistical and research uses of personal health data to advance medical research and health care quality. When surveyed, the answers provided seem to depend on how informed the survey participants are about research uses of data and whether the survey is about an abstract, hypothetical situation or is focussed on a particular research centre, bio bank or government ministry, with community recognition and trust. Further there may be cultural or historical reasons for differences in views, as well as differences resulting from how the survey questions are phrased.

In Europe, a survey conducted in 2012 indicated that 74% of Europeans view health data about them as personal data (Lusoli et al., 2012). Among this group, the majority (76%) felt that their approval should be given in all cases when personal data of any type were to be collected and processed and 74% would be concerned if there were unannounced uses of their personal health data for purposes for which it had not originally been collected. At the same time, most expressed trust in national authorities (73%) and health care providers (86%) to protect their personal identities within health data.

A Eurobarometer survey indicated that only 34% of Europeans had heard of bio-banks prior to the survey (Eurobarometer, 2010). Awareness varied a great deal from 80% in Iceland and 75% in Sweden to 19% in Portugal, 18% in Austria and 15% in Turkey. The lack of knowledge about biobanks renders the rest of the poll quite speculative. Nonetheless, the poll reported that 46% of Europeans would be willing to provide information to a bio bank. Further, most expressed the view that researchers using data stored in biobanks should ask their consent for each new research project, from 51% in Denmark to 85% in Greece. The survey, however, did not provide any information to participants about the impact of this requirement on the validity of research studies nor on themselves in terms of their future response burden.

A recent study of the general public in the United Kingdom involved a quantitative online survey and focus groups (Lewis et al., 2013). It solicited views on willingness to donate biological samples for use in biomedical research including organs, tissues, fluids and genetic material (DNA) obtained from medical procedures or post mortem. The survey found that 87% of persons would donate samples for research. Participants were asked if they would be willing to have their bio sample data linked to their clinical records and lifestyle information if the data were de-identified. Sixty-eight per cent of focus group participants said they would, 22% said they would not and 10% did not know. Among survey respondents, 82% would, 12% would not and 6% did not know.

In Canada, the Canadian Institute for Health Information and Canada Health Infoway commissioned a public opinion poll in 2012 regarding Canadians views about the use of data extracted from electronic clinical records for statistics and research (CIHI, 2012). The poll results indicated that Canadians' understanding of an electronic health record as a tool that goes beyond record keeping has increased since 2010. Most Canadians believe that the analysis of data from electronic health records will have an impact on the health care system and patient care. There remains a significant share of Canadians that are very concerned (14%) or somewhat concerned (22%) about the secondary use of their individual patient data. The proportion somewhat concerned about secondary data use declined to 15% if the data are de-identified. Over 70% of Canadians are comfortable with the secondary use of data by government, health care organisations, universities and statistical offices. In contrast, only 29% of Canadians are comfortable with the secondary use of data by commercial interests, such as drug or insurance companies. A strong majority of Canadians would like to give their consent before identifiable personal health data was used for secondary purposes and a similarly strong majority would say yes to health system use of their identifiable data, such as for research to track how well treatments are working.

Three US surveys were conducted with populations living near a bio bank. In the first, residents of a community with an established bio bank were very favourable toward it (Simon et al., 2011). Eighty-nine per cent were confident that their identities were protected when data were used for research; and 90% strongly agreed to keep the opt-in consent model used by this bio bank. In this opt-in model, participants consent to future uses of their data. In the second, in a community where a bio bank did not yet exist, 67% of respondents preferred an opt-in consent to future research with their data over an opt-out consent approach (Brothers et al., 2011). In this survey, broad, research-unspecific consent was preferred (54%) over categorical (21%) and study-specific (21%) consent models for purposes of approving future research use. In the third survey, an HMO planning a bio bank asked patients visiting clinics to read an informational brochure and an informed consent form. They were then asked to complete a survey. In this survey, 69% indicated their willingness to contribute to the bio bank using an opt-in consent model. Of those with concerns, 35% indicated information security was the reason (Rahm et al., 2013).

Another view comes from the experience of national health examination surveys. These surveys, administered by a central governmental authority, ask a random sample of citizens to volunteer to provide data and biological samples for future research studies using an opt-in consent methodology. The first administration of the Canadian Health Measures Survey in 2010 resulted in 88% of screened households having an individual participate in a voluntary health survey and among them, 85% also attended a health clinic where biological samples and physical measures were taken (Statistics Canada, 2011). In the 2011 administration of the US National Health and Nutrition Examination Survey, 73% of screened households participated in a voluntary health survey and 69% participated in a health clinic where biological samples were taken (NCHS, 2013).

In summary, it seems that citizen views depend upon the level of abstraction of the data collection, from hypothetical scenarios to more concrete questions involving known and trusted health care providers or government organisations. Citizen views are also influenced by their personal awareness of the issues discussed, such as biobanks and the use of biomedical information in research studies. Further, it also seems evident that even among well-informed populations there are always a certain percentage of people that would prefer that their personal data are not used for research.

Public communication: Lesson’s learned from the UK Health and Social Care Information Centre

In many ways, as has been highlighted in this chapter, the United Kingdom has a strong policy toward openness and transparency about data and data access. Nonetheless, and perhaps because of this openness, it also has experienced difficulties that provide important lessons-learned about the necessity of strong public communications that must accompany openness and transparency initiatives.

Strategic plans, public consultation and the launch of care.data

In the United Kingdom, the health and care system in England has undergone a major reorganisation over the last two years with a number of new organisations established from April 2013. The Department of Health's role is now as steward of the new system and to set the overall strategic direction for health, public health and adult social care in England. In particular from April 2013 the Health and Care Information Centre (HSCIC) was established with new powers and responsibilities in relation to data. In common with a number of other government departments the Department of Health has a transparency board, the Health and Social Care Transparency Panel (HSCTP) whose role it is to give advice on making data more transparent and available.

There was a public consultation on the information strategy for health including the approach to developing information infrastructure for health and social care. This public consultation took place between October 2010 and January 2011. The consultation received 742 responses from all stakeholder groups, for example from clinicians and the voluntary sector. It was a broad response that led to the development and publication of the strategy in May 2012 by the Department of Health entitled "The Power of Information: Putting All of Us in Control of the Health Information We Need". This report sets out a ten-year vision for an information system that collects clinical and patient-level data once that can then be used, reused and shared in accordance with strict rules to protect patient privacy. The goal is to share information within the NHS and Social Care that supports delivering care, that also provides, securely, data for research and statistics, and that enables patients to access their own records to empower them and to help them to manage long-term conditions and to monitor their own care.

The strategy includes data linkage for research purposes, where data are linked and then anonymised or pseudonymised before they are provided to researchers under clear information governance requirements. The strategy aims toward data that will be more effective for planning because it will be timelier (closer to real-time). The objective is to increase the sharing and use of data including to collect data once at the point of care and to create a national repository for the data. The programme is called care.data.

Care.data will provide a modern data system for the NHS, to provide patient's with access to their own records, to provide services and data for research, to support policy bodies and research organisations e.g. King's Fund and the Nuffield Trust. The data held within HSCIC from care.data will be at the level of patients or health care events and be linkable across different data sources, pseudonymised and then made available for use subject to strict information governance rules. Better IT systems will enable data that help the NHS to manage demographic pressures on health and social care budgets by providing the data needed to make health care more efficient.

England is in the process of developing this new approach. At present, there is a lot of health data available, particularly held within the Health and Social Care Information Centre (HSCIC) which is the main custodian of England's health data. HSCIC offers a data linkage service to approved applicants. Approved applicants pay a fee for the linkage service which covers the cost of conducting the linkages. While HSCIC was undertaking linkages in the past, it was in September 2012 that the services were announced and certain data linkages became undertaken on a routine basis, such as a monthly linkage of mental health patient data to hospital episode statistics. Applicants for data linkage services must explain the data they want to use and how they intend to use it. The HSCIC also provides data linkage services to the Clinical Practice Research Datalink (CPRD) which began providing services in March 2013. CPRD is an advisory service that guides researchers preparing research proposals for approval including getting access to linked patient-level data (pseudonymised).

A review of health data confidentiality protections and information governance was conducted by Dame Fiona Caldicott and published in 2013. This was undertaken, in part, to address concerns that clinicians and health and care professionals were reluctant to share data even where it would be for the benefit or safety of patients, because of concerns about respecting legal requirements. The review looked at risks of both sharing and not sharing data. The review found that most patients expect the sharing of data for their personal benefit, such as for better care but there is nervousness about the use of confidential data by government.

Care.data: Lessons learned about transparency and public engagement

In the winter of 2014, a media story raised public concern about the potential for commercial interests to purchase a dataset containing patient-level records from the HSCIC. Concerns that were raised centred on the transparency of the care.data initiative. Particularly, there was criticism about the adequacy of public consultation and public communication about the initiative and questions raised about the safeguards to protect privacy and to ensure data uses would be in the public interest.

Prior to the media story, NHS England, with the HSCIC, had undertaken a public awareness raising campaign about the care.data initiative by sending a leaflet to all households to explain the data they have and how it is being used and the benefits of data use and the protections of information privacy. The leaflet explained that individuals may opt-out of national data repositories. There was also a You-tube video that was included on the NHS website conveying the same messages. For fiscal reasons, there was no advertising budget for the campaign.

After the media story, the initiative was put on hold for six months to raise awareness, listen and act on the views of patients and key stakeholders, and to discuss both the benefits and risks involved (NHS, 2015). Issues for public discussion include ensuring the public are aware of the initiative, have a clear understanding of the issues and understand their right to opt-out of data repositories.

The legislative framework for health information governance was also strengthened. The *Care Act* was introduced in May 2014. This legislation restricts the sharing and analysis of personal health data to uses that have a public benefit to health or social care or to health promotion. The law also strengthens data governance by requiring the Health Research Authority to establish an independent body to advise the Secretary of State, the Health Research Authority and the Health and Social Care Information Centre regarding the processing or sharing of personal health data.

There will be a pilot phase of care.data where a small number of GP practices will submit patient data to the health and social care information centre in order to pilot test ways of supporting GPs to ensure patients are informed of the purposes of this data sharing, its safeguards and how they can object/opt out. The Independent Information Governance Oversight Panel (IIGOP), chaired by Dame Fiona Caldicott, will also advise the HSCIC on the first phase of implementation of care.data.

Approaches to engaging with stakeholders and the public about the processing of personal health data

The Secure Anonymised Information Linkage (SAIL) project in the UK Wales is an example of a multi-faceted effort to engage with stakeholders in the use of personal health data. A consumer panel was established with members of the public who are involved in advising either hospital or social services. These panellists have now been involved in SAIL for two years and are very knowledgeable about the data governance of SAIL. Very recently, the consumer panel has asked SAIL to improve its transparency to the public by providing summaries of approved research projects on its website. SAIL is communicating with researchers in advance of the implementation of this request. The consumer panel plays an important role in educating the broader community about SAIL as they speak about SAIL to their constituencies. SAIL also has research fellows specialised in public engagement and data governance and their work improves public outreach and involvement. Larger projects within SAIL each have established steering groups that include members of the public or patients. SAIL hosts the UK MS registry and patients with this condition advise on the development, management and use of the registry. SAIL approaches GP practices individually to secure their participation in the project. SAIL has an information package and a Q&A for General Practitioners (GPs) considering signing up. GPs are welcome to participate in research projects within SAIL as a research leader or member of a research team. There is a newsletter sent to GPs that provides feedback to GPs about SAIL projects with GP data. GP practice codes are masked within SAIL to assure GPs that SAIL is not engaged in performance management activities. Upon request, GPs can receive results for their own practice, but with de-identified patient data.

In Scotland, Farr@Scotland, which is part of the Farr Institute, is bringing together people working in accredited safe havens, public authorities and health researchers to discuss standardising processes for data governance. There has been opportunity to explain the data needed for research and the processes for anonymisation and data security that keep data used for research safe.

Farr@Scotland is also conducting a programme of public engagement activities in order to raise awareness of the ways that researchers use data and the types of research that are conducted, whilst also finding out what people think about this and feeding back public views into decision-making processes within Farr@Scotland and the wider UK Farr community. A public panel of 20-30 lay representatives has been formed. The panel advises Farr@Scotland on how best to inform the public about Farr and how best to inform them about the benefits of health research and how the data is protected. Challenges for successful public communication include that the public cannot see how research with data has a direct impact on their health, and many people believe that the level of data access Farr is working to achieve already exists.

Key elements of data governance that promote openness and transparency

This chapter highlights progress in the openness and transparency of governments regarding the collection and use of personal health data and emphasises the importance of not only having strong and effective data governance mechanisms but also ensuring that public communications about data governance are effective and reach their intended audience. While there is no doubt that communicating about uses of and safeguards surrounding personal health data is challenging, transparency is essential to maintaining public trust and confidence in government.

The Advisory Panel of Experts on Health Information Infrastructure identified the following key elements of the governance of health data that promote openness and transparency:

3. The public are consulted upon and informed about the collection and processing of personal health data

Public engagement

a) Includes regular, clear and transparent communication with the public about the collection and processing of personal health datasets including the benefits of the processing, the risks of the processing and the risk mitigations.

b) Includes public information, such as a website, that describes personal health datasets at a national level, including the content of the datasets and the dataset custodians.

c) Includes public information, such as a website, that describes applications for approval of the processing of national personal health datasets, including dataset linkages, as well as approval decisions.

References

CDC – Center for Disease Control (2015), NCHS Research Data Center, Center for Disease Control, United States, www.cdc.gov/rdc/, accessed 16 February 2015.

CIHI – Canadian Institute for Health Information (2014), "Data De-identification Best Practice Guidelines", Canadian Institute for Health Information, see www.ehealthinformation.ca/wp-content/uploads/2014/08/2011-Best-Practice-Guidelines-for-Managing-the-Disclosure-of-De-Identificatied-Health-Info.pdf, accessed 30 July 2014.

CIHI (2012), "Public Opinion Research: Canadian Views on Electronic Health Records", Canadian Institute for Health Information, www.cihi.ca/CIHI-ext-portal/internet/EN/TabbedContent/about+cihi/corporate+strategies/health+system+use/cihi011257 accessed 30 July 2014.

DHHS – Department of Health and Human Services United States (2014), http://healthdata.gov/blog/health-data-initiative-strategy-execution-plan-released-and-ready-feedback, accessed 29 July 2014.

Eurobarometer (2010), Biotechnology, European Commission, http://ec.europa.eu/public_opinion/archives/ebs/ebs_341_en.pdf.

Lewis, C. et al. (2013), "Public Views on the Donation and Use of Human Biological Samples in Biomedical Research: A Mixed Methods Study", *British Medical Journal Open*, Vol. 3, No. 8, p. e003056.

Lusoli, W. et al. (2012), "Pan-European Survey of Practices, Attitudes and Policy Preferences as Regards Personal Identity Data Management", JRC Scientific and Policy Reports, European Commission.

National Centre for Health Statistics, National Health and Nutrition Examination Survey, www.cdc.gov/nchs/nhanes/response_rates_cps.htm accessed 8 October 2013.

NHS – National Health Service (2015), "The Care.data Programme – Collecting Information for the Health of the Nation", National Health Service, England, www.england.nhs.uk/ourwork/tsd/care-data/ accessed 15 February 2015.

Rahm, A.K. et al. (2013), "Biobanking for Research: A Survey of Patient Population Attitudes and Understanding", *Journal of Community Genetics*, Vol. 4, No. 4, pp. 445-450, http://dx.doi.org/10.1007/s12687-013-0146-0.

Simon, C.M., J.L. L'Heureux and B. Zimmerman (2011), "Active Consent but Not Too Active: Public Perspectives on Biobank Consent Models", *General Medicine*, Vol. 13, No. 9, pp. 821-831.

Statistics Canada (2011), "Canadian Health Measures Survey Data Users Guide", Ottawa.

THL – National Institute for Health and Welfare, Finland (2014), www.THL.fi, accessed 14 August 2014.

United Kingdom (2012), "Open Data White Paper: Unleashing the Potential", http://data.gov.uk/sites/default/files/Open_data_White_Paper.pdf, accessed 5 August 2014.

Chapter 5

Concentrating and strengthening national health data processing

This chapter discusses current centralisation of national dataset processing as well as the implications of centralisation for the protection of health data privacy and the accessibility of data for approved research and statistics. It also discusses the introduction of accreditation or certification of data processors.

Highlights

Seventy per cent or more of the key datasets of personal health information are held by a single organisation in ten countries. Countries with concentrated custodianship of national data have several distinct advantages. These custodians can conduct data linkage projects without entering into negotiations and data sharing agreements with other data holders. Thus they are more likely to have regular programmes to monitor health and health care quality and performance that are based on data following the pathway of care. They are also more likely to be resourced sufficiently to develop efficient data processing and to provide timely and high quality services to external data users. Most custodians that provide services to external data users have developed mechanisms to recover some of the costs associated with these services.

Concentration can be argued as being a risk to data privacy and confidentiality protection. However, with good governance mechanisms assured through accreditation or certification, the risks from concentration can be managed. An accreditation or certification process can narrow the number of processors to only those who meet the country's highest standards for data privacy and security protection. Further, follow-up audits can ensure that these standards are maintained. Accreditation of data processors is under consideration in the United Kingdom (England) and has been introduced in the United Kingdom (Scotland) and in Australia.

Whether by policy or by default, some countries have concentrated the collection and processing of key national personal health datasets. As a result, they have distinct advantages in the further development of these data for statistics and research, in the undertaking of approved data linkages studies and in organising and improving secure access to data for external researchers and public bodies. This chapter discusses current concentration of national dataset processing and the introduction of accreditation or certification for central data processors and the implications for the protection of health data privacy and the accessibility of data for approved research and statistics.

Concentration of national health datasets

The greatest concentration of national health datasets is in Switzerland and Turkey where all key national datasets are in the custody of the same organisation (Table 5.1). Ninety per cent of national datasets are within one organisation in Iceland and Japan. Other countries with a high proportion of national datasets concentrated in one custodian are the United Kingdom (Scotland) at 78% followed by Denmark (75%), New Zealand (75%), the United States (73%), the Czech Republic (71%), and Sweden (70%).

Data linkages are concentrated in many countries

Half of the countries in this study regularly conduct data linkages of all of their key national datasets within a single organisation (Table 5.1). Of course, as was discussed in Chapter 2, some countries have national health datasets for key components of the continuum of health care and others have more limited national data. Table 5.2 presents how countries are positioned regarding the number of key national datasets that are regularly included in data linkage projects and the number of organisations usually linking the data. Only four countries are regularly linking eight or more key national health datasets within a single organisation [United Kingdom (Scotland and Wales), Singapore and Iceland].

Table 5.1. Proportion of national datasets in the custody of and linked by the same organisation

	% of key national datasets in the custody of the same organisation	% of datasets usually linked by the same organisation
Canada	60%	71%
Czech Republic	71%	100%
Denmark	75%	100%
Finland	64%	67%
Iceland	91%	100%
Ireland	25%	33%
Israel	67%	75%
Italy	67%	100%
Japan	90%	0%
Korea	55%	75%
New Zealand	75%	100%
Netherlands	33%	100%
Norway	36%	100%
Singapore	60%	100%
Spain	33%	50%
Sweden	73%	64%
Switzerland	100%	100%
Turkey	100%	100%
United States	73%	50%
UK England	56%	67%
UK Scotland	78%	100%
UK Wales	33%	100%

Source: Author’s own calculations based on the results of this study.

Table 5.2. Number of organisations and datasets involved in linkage of key national health datasets

	Largest number of key health datasets regularly linked by the same organisation	Number of organisations regularly linking key national datasets
UK (Scotland)	10+	1
UK (Wales)	10+	1
Singapore	10	2
Iceland	8	1
Denmark	7	1
Switzerland	7	2
Sweden	7	3
Netherlands	6	2
Finland	6	3
Israel	6	3
Korea	6	3
UK (England)	6	3
New Zealand	6	4+
Czech Republic	5	1
Canada	5	2
Norway	4	3
United States	3	4+
Italy	2	2
Spain	2	4+
Ireland	1	3
Japan	0	1
Turkey	0	1

Source: Author’s own calculations based on the results of this study.

Data processing centres

A few countries provided additional information about efforts to concentrate data processing and access.

In Japan, the Bureau of Health Insurance holds insurance claims and data regarding health check-ups required by employers. The insurance claims and health check-up datasets are the first national datasets that have been made accessible to researchers with approved projects in Japan. The health insurance claims data covers 90-95% of all health care treatments delivered in Japan. The data have been collected from 2009 onward and, as a result, there is five years of history of health care experiences that can be analysed. The data are big with an estimated 130 million claims per month.

In Denmark, the Statens Serum Institut (SSI) was given responsibility for all of the major national registries related to health in 2012. This reorganisation enabled the major national registries related to health to be consolidated within one organisation. In so doing, there should be more synergies from the use of health registry data including an improvement in the feasibility of national data linkages, such as, for example, primary health care to secondary care. The SSI will provide access to data for other ministries as well as conduct data linkages that those ministries may require. The SSI also provides data linkage services to researchers and will link an external dataset to the SSI registries for an approved project. Other organisations including Statistics Denmark have health data and can conduct data linkages.

Canada is a federation where individual provinces and territories have the jurisdiction to manage and deliver health care. The Canadian Institute for Health Information (CIHI) was formed following a critical review of the state of national health information. The Canadian provinces and territories contribute financially to CIHI, sit on its board of directors and develop, via consensus, agreement about data collections and the data standards necessary for comparable and high quality data at a national level. Consensus building of data standards ensures there is a common core of data, as well as provides flexibility for individual provinces to include elements that are only of interest to them. The model has been working. In order to receive identifiable personal health data from the province of Ontario, CIHI was granted the status of prescribed entity within Ontario health data privacy law. Thus CIHI's data privacy and security framework is consistent with the privacy-by-design requirements of the Ontario Data Privacy Commissioner's Office. With both a strong data privacy protection framework and a process to develop consistent standards for data reporting, then there is opportunity for personal health data to be used to provide information about quality and performance at the national level within this federated country. In Canada there is also no policy to concentrate data processing or access services; however most processing takes place within the Canadian Centre for Health Information with some processing also within Statistics Canada and other federal health ministries and agencies.

The Health and Social Care Information Centre (HSCIC) in England was amalgamated with Connecting for Health in April 2013. The HSCIC became recognised in law as the safe haven for the collection, processing, analysis and dissemination of data about the health and social care system in England. Included in its work is the production and publication of indicators and information, as well as work to develop the information infrastructure for the health and social care system. There is a large programme of work on the development of electronic health records including content/standards and networking services. The work of the HSCIC covers England only. The HSCIC run a data linkage service and are an integral part of the care.data initiative. Under the care.data initiative there will be new datasets of record-level data developed that can be linked for approved projects in order to provide more information about patient's health and social care journeys. Care.data will bring together primary health care data to be linked with secondary care data. Within the HSCIC in England, the Director of Information Assurance is the head of information governance and ensures that practices in HSCIS are in line with legal requirements. Supporting the director is a governance team including the Caldecott Guardian and the Head of Statistics. The Caldecott Guardian is the senior person responsible for the safe handling of personal data within the HSCIC. The *Care Act*, introduced in 2014, established an independent review body that will advise the HSCIC on projects involving the processing of personal health data.

The Secure Anonymised Information Linkage project (SAIL) was started in Wales by a government grant in order to enable data routinely collected from health care provision and other government service delivery to be anonymised and linked by a trusted third party, which is a National Health Service organisation (SAIL, 2014). Overall, SAIL includes national data that cover the whole population of Wales (3 million people) from 1990 onward and are both broad and deep. Datasets included within SAIL are national hospital in-patients and out-patients data, cancer registry data, cancer screening data, and emergency visits data. Also included in SAIL are data on primary health care visits extracted from GP electronic clinical record systems. For these data, SAIL negotiates with GP practices and about 40% of practices in Wales are participating. Health boards in Wales are now supporting SAIL and this support is expected to increase GP participation in the future. SAIL also includes contextual data for health studies, such as education and housing data.

In Iceland, there is no policy to concentrate data processing; however, most processing takes place within the Directorate of Health who is responsible for and maintains all national health registries. This concentration helps because it is necessary to have staff familiar with the content of databases and how they are coded to undertake requests for data access and linkages. The Directorate is building a national data warehouse and is now receiving patient data from hospitals in real time regarding both admissions and discharges and out-patient encounters. Prior to this development, hospital data were submitted to the Directorate annually. Under the new system, all hospitals and physicians are using the same electronic health record system which is improving the timeliness, quality and comparability of the data.

Data processing at the national level is concentrated in Switzerland within the Federal Statistical Organisation. By law, the federal FSO is the only organisation authorised to link the data that it holds.

Accreditation or certification of data processors

There have been recent efforts in the United Kingdom and Australia to introduce an accreditation process for organisations wishing to process health data. Accreditation provides a means to establish detailed data governance criteria that accredited organisations must meet, to independently verify that the requirements are satisfied before granting accreditation status and to audit organisations that are accredited for compliance.

Following a public consultation, an independent Information Governance Review Panel in the United Kingdom proposed that the linkage of personal health data or of data with a high re-identification risk be conducted within a well-governed, independently scrutinised and accredited environment called an accredited safe haven (Information Governance Review Panel, 2013). The report recommended the following set of standards for accreditation as a safe haven and recommended that any accredited organisation be subject to an independent external audit for compliance:

1. attributing explicit responsibility for authorising and overseeing the anonymisation process e.g. through a Senior Information Risk Officer
2. appropriate techniques for de-identification of data, the use of 'privacy enhancing technologies' and re-identification risk management
3. the use of "fair processing notices"
4. a published register of data flowing into or out of the safe haven including a register of all datasets held
5. robust governance arrangements that include, but are not limited to, policies on ethics, technical competence, publication, limited disclosure/access, regular review process and a business continuity plan including disaster recovery
6. clear conditions for hosting researchers and other investigators who wish to use the safe haven
7. clear operational control including human resources procedures for information governance, use of role-based access controls, confidentiality clauses in job descriptions, effective education and training and contracts
8. achieving a standard for information security commensurate with ISO2700161 and the Information Governance Toolkit
9. clear policies for the proportionate use of data including competency at undertaking privacy impact assessments and risk and benefit analysis
10. standards that are auditable
11. a standard template for data sharing agreements and other contracts that conforms to legal and statutory processes
12. appropriate knowledge management including awareness of any changes in the law and a joined up approach with others working in the same domain
13. explicit standard timescales for keeping datasets including those that have been linked, which should be able to support both cohort studies and simple "one-off" requests for linkage.

Within the United Kingdom, England is considering introducing Accredited Safe Havens. The HSCIC already has features of a safe haven and there is discussion about the number of safe havens that could be created. The position of the ICO is that the number is not as important as is the strength of the accreditation required to become a safe haven. The standards that an organisation should have to meet to be accredited should be high and there should be vigilance to ensure that there is no abuse and strong sanctions in the event of any abuse.

Accredited safe havens in Scotland

Within Scotland there are currently five health-related accredited safe havens. All of these sit within the NHS but four were developed collaboratively between Health Boards and Universities as part of NHS Research Scotland (Aberdeen, Dundee Edinburgh and Glasgow); the fifth, at NHS National Services Scotland, was developed as part of the Scottish Health Informatics Programme. Each safe haven has evolved in a slightly different way to meet the broad requirement of providing a safe environment to process health related data for research purposes. The regional datasets can include deep phenotypic data (biological, genetic etc.) whereas the national datasets are often more general but cover a larger population. All of these safe havens provide data linkage services and access to de-identified health microdata to approved researchers through a secure real-time remote data access system. Key features of the Scottish Health Informatics Programme (SHIP) accredited safe havens are as follows (SHIP, 2014):

1. The Safe Haven will provide a secure environment for the linkage, storage and analysis of personal data.
2. The Safe Haven will hold datasets and ensures that only approved researchers can gain access.
3. Researchers access the data held within the Safe Haven via a dumb terminal in a secure access facility. The dumb terminals will be configured so that the researcher cannot download or remove any of the data or outputs held at the Safe Haven.
4. Analytical software will be available within the Safe Haven for use by researchers.
5. A dedicated file space will be provided for the researcher to store their outputs pending release by the Safe Haven.
6. Safe Havens will carry out statistical disclosure control on outputs to prevent accidental disclosure of identifiable information.
7. There will be penalties for anyone who abuses personal data. Researchers will be bound by a strict code, which prohibits disclosure of any personal identifying information.

Within Scotland, each Safe Haven adheres to the *Data Protection Act*, Caldicott Principles (where required), Data Sharing Agreements, Governance Agreements, Ethics Approvals and other relevant agreements. Historically a new set of specific agreements were developed for each dataset to be held or processed by a safe haven, each dataset to be transferred into and out of a safe haven; and each data linkage project carried out by a safe haven.

The agreements were not standardised, so each time, the process to develop agreements was extremely time consuming, required significant resource and the resulting agreements often differed across projects and Safe Havens.

To address this, the Farr Institute @ Scotland, part of the United Kingdom Farr Institute, is developing a Federated Model of Safe Havens. A Safe Haven Charter is being developed that would provide harmonised agreements and principles under which an accredited Safe Haven would adhere; thus streamlining and standardising the process of storing, processing and linking data. The aim is to ensure that procedures and practices are transparent and subject to independent audit. Standardising processes will allow exchange of data among safe havens for collaborative projects with the knowledge that the same data security and privacy standards are being met in any safe haven. The Safe Haven standardised processes should meet the expectations of data controllers and ethics committees in terms of data privacy protection and data security and develop a well-regulated and supportive research environment in Scotland.

The five accredited safe havens are coming to an agreement that when there are data linkage projects involving multiple data custodians, identifiable data will be sent to an independent indexing service. The indexing service will clean the identifiers, remove the unique identifiers and assign a random study number. A trusted third party will perform data linkage and data de-identification and then provide the safe haven with a de-identified linked dataset for analysis. Researchers have to analyse the data within a Safe Haven environment via a remote log in or from within secure physical premises. Only aggregate results are allowed out of the Safe Haven environment. Statistical disclosure is recognised.

This method enables data custodians to be confident that no one organisation will see linked identifiable data; that the data will never leave the Safe Haven environment; and that data will still be under their control. Thus, the method has persuaded data custodians to contribute data to the safe havens. However, there are several challenges resulting from this approach:

- The work to index, extract, link and provide a dataset has to be completely redone every time an extract is required. This is time consuming and costly.
- The dataset cannot be re-used by other research groups and thus replication of findings is not possible.
- The researchers' final datasets used for their analysis are not archived and each research group may clean and process their data in a different way, contributing further to the inability to replicate findings.
- As different organisations are involved at each step, it is difficult to trace back any errors in the data.

Accredited integrating authorities in Australia

In Australia, an authorised and accredited integrating authority must be identified for each statistical data integration proposal involving national data (Australia, 2014a). The integrating authority will ensure appropriate governance is in place for data integration and data linkage projects including:

1. using an open approval process
2. documenting the proposal
3. considering the privacy impacts
4. examining the expected costs and benefits of the proposal
5. considering the access arrangements and dissemination plans.

The integrating authority will be responsible for the ongoing management of the integrated data, ensuring it is kept secure, confidential and fit for the purposes for which it was approved.

Criteria for accreditation as an integration authority have been established. All applicants must be subject to the provisions of the *Privacy Act*. The Australian Institute of Health and Welfare was designated an accredited integrating authority in 2012. The AIHW application for accreditation is publicly available and describes how the AIHW meets the following criteria (Australia, 2014b):

1. ability to ensure secure data management
2. ensure that information that is likely to enable identification of individuals or organisations is not disclosed to external users.
3. availability of appropriate skills.
4. appropriate technical capability
5. lack of conflict of interest.
6. culture and values that ensure protection of confidential information and support the use of data as a strategic resource.
7. transparency of operation.
8. existence of an appropriate governance and institutional framework.

Processing data access requests and recovering their costs

Providing services to researchers, such as advising upon data access requests, approving requests and processing approved data requests, costs data custodians in time and resources. Most countries that are more open about providing services to researchers also have developed mechanisms to recover some of these costs. Nonetheless, a number of countries expressed concerns about insufficiencies in the availability of staff or in financial resources to keep up with requests for access to data.

It is worth noting that there is a significant difference in resource expenditures between countries where there is a unique patient identifying number that is widely used and accurately captured within key national health datasets and those where there is a need to rely upon a set of potential identifying variables in order to establish a dataset linkage for statistics and research. In the former, providing data linkage services can be a routinised and automated process requiring few resources. In the latter, data linkage services require skilled statisticians and considerable time to execute accurately.

Countries handling a high volume of data processing requests each year

Denmark and New Zealand have developed their health information infrastructure to the point where they are currently supporting a high volume of requests for data processing services every year and the United Kingdom (England) is preparing for such volumes in the future.

In Denmark, there is no clear tracking of the volume of requests from municipalities, regions and hospitals. For applications from external researchers, there are about 2000 applications per year. About 60% are requests for access to a single registry and 40% are requests for data linkages among registries. Requests are received from public authorities, academics and non-profit researchers and from the commercial sector. There are still only a limited number of requests for multi-country projects. In Denmark, in general it takes between 1 and 15 hours to process a request depending on its complexity. The average time is about 3-4 hours of work per request. More complex requests may require the involvement of additional staff members, such as the registry manager and, consequently, take longer to process. In Denmark, the time it takes to process a request is charged using an hourly rate of about EUR 175 per hour. In Denmark, at the SSI, to process requests for access to registries and registry linkages there are 8-9 dedicated staff plus students for a total of about 14 people.

In New Zealand, the Ministry of Health estimates that there are about 2 500 requests for data annually, and among them 50-100 requests require data linkage. The vast majority of requests are from government or academic researchers. Requests from commercial interests or foreign entities are rare. In New Zealand, in total there are between 25 and 30 employees involved in conducting data linkages. The staff time taken to fulfil a request is recovered from the client. For many data linkages the process is very efficiently undertaken as it is a direct linkage involving a consistently applied health number. New Zealand notes that data de-identification is an automated process and is not resource intensive, but there is a lack of skilled personnel capable of assisting in data linkage and de-identification efforts.

The Health and Social Care Information Centre in England, introduced in 2013, is gearing up to support a high volume of data processing requests. In the United Kingdom, the *Freedom of Information Act* allows the public to ask for custom analysis with results returned to them within a limited time window of 20 days. The HSCIC is able to mostly meet all of these FOI requests within the time limit. In cases where a data requestor requires an extract with a high re-identification risk then the process is longer, including the time for the external Research Ethics approval and then the development of a data sharing agreement. The process varies from a few weeks to a few months for data linkage requests. In England, the fees charged by the HSCIC are provided on their website. In general, the price ranges from GBP 1 000 for a data extract to tens of thousands for a customised data linkage service. The legislation establishing the HSCIC does not allow the HSCIC to have different prices for different data requestors. As a result, there is no lower price offered to students. The HSCIC signalled that data linkage systems take considerable time and resources to set up.

Countries with a moderate volume of data processing requests annually

A group of OECD countries have well-established data processing services meeting a more moderate number of data processing requests each year (Finland, United States, Canada, Iceland and Norway).

In Finland, there are approximately 150 requests for access to health and/or social care microdata each year and over half of these requests are for a data linkage. Finland provides assistance to researchers to prepare their application and, as a result virtually all of the applications are approved. Most applicants from outside THL are clinicians or social scientists from within hospitals or academic centres. Clinician researchers do request data linkages for clinical trial cohort follow-up. PhD students also frequently request data linkages and access to data. There are incentives in Finland for clinicians to engage in data analysis and to be able to share their time between research projects and clinical work. There is some funding available and they can also request time off from clinical work to conduct research. THL makes an effort to encourage clinicians and health professionals to engage in data analysis by, for example, speaking at practitioners annual meetings to describe the data available and how it can be analysed to explore research questions. An example was a presentation made to midwives who were initially resistant to data collection and use but then became more interested in the data as they learned more about it. There are now both midwives and nurses who have received PhDs for projects using THL registry data.

In Finland, the staff time taken to fulfil the request is recovered from the researcher at a rate of EUR 120-250 per hour. THL in Finland estimates that it takes between one day to three months to process a data access request and the duration depends on the applied special enactment.

In the United States, the NCHS team working with applicants for access to microdata have two different types of data requestors, those requesting a merge of microdata with socio-economic data describing small geographic areas (contextual area-based data) and researchers requesting a data linkage or access to data that are already linked. There is an average of 170 new requests processed annually and about 700 projects underway annually. About 30 of the 170 requests are for record linkages. Most of the requestors have received a scientific grant from the US National Institutes of Health. The NCHS works with the researchers preparing applications for access to data to help them to define their study and their access request in a way that it can be approved. As a result 95% of applications are approved. Among the applications received, only 1% are referred to the NCHS Ethics Review Board for approval because they raise concerns that were not addressed at the working level. In the United States, a request submitted by a researcher with past experience in preparing an application for data access can be approved and processed in as little as a day. The process can be time consuming for an applicant that is unfamiliar with the requirements for approval or who has a complex request and can take months or even years in some cases.

In the United States, there are 17 employees working in the Research Data Centers. Counting these employees plus the employees involved in preparing the data for researchers, as well as those involved in working with researchers on their applications and reviewing and approving applications, there are about 35 full time employees involved. In the United States, the costs of project approval, preparing data and providing access to data are recovered from approved applicants for access to NCHS data. The daily rate for management services to prepare the data, guide the researcher in their application and review and approve proposals is USD 750 per day. The daily rate for access to data within the NCHS Research Data Centers is USD 300. The Andre secure data access system is provided at a monthly rate of USD 750. If NCHS analysts are required to conduct some of the analysis their time is recovered at a daily rate of USD 750. All of these rates are posted on the NCHS website. There are considerable fixed costs involved with the infrastructure and maintenance of the RDCs, including the management of the paperwork and the IT infrastructure that would be equally high regardless of country size or size of the research community.

There are approximately 30 approved and processed data linkage projects conducted each year by the Directorate of Health in Iceland. Most of these projects are for scientific research and were requested by academic or non-profit researchers. This number excludes data linkages the Directorate leads to improve the quality of its own registries. There are many more requests that are also processed for data tabulations. The elapsed time to conduct a linkage project is not measured in Iceland. It depends on the quality and complexity of the application. Some applications must be revised and re-submitted if they raise concerns and the process takes months to resolve. The data linkages are carried out by the Directorate in the order they are received following approval. In Iceland, the bioethics committee meets twice per month which supports timely responses. Further, the timeliness of the process is improved when the researcher submits an application to the Bioethics Committee, the Data Protection Authority and the data custodian at the same time. The DPA will then be able to begin its review of the application earlier and, as a result, conclude it more quickly after receiving the advice of the Bioethics Committee and the approval of the data custodian. Iceland reported concerns with insufficient staff and financial resources to undertake data linkages.

The Norwegian Patient Register has approximately 150 requests yearly for access to microdata and other institutions, such as the Norwegian Institute of Public Health, also process data access requests.

In Canada, the majority (90%) of data sharing activity within CIHI involves providing data that have been cleaned, validated and enhanced with new derived variables (such as case mix) or calculated indicators (such as hospital standardised mortality rates) back to the original data owners who provided the data to CIHI, such as individual hospitals and health care facilities in some provinces, or provincial organisations. Requests for detailed record level data or data linkages are in the low 100s annually. Only 10% of requests involve third parties, such as academic researchers.

In Canada, CIHI's commitment to the timely fulfilment of data requests as well as cost-recovery rates for services is all posted on their web-site. CIHI aims to fulfil simple requests within ten days, requests for record-level data from individual datasets within 20 days. Requests for data linkages would take a bit longer due to the time for the project approval process (3-5 months). The tool for automated tracking of data requests and approvals has helped to speed up the process. In particular, if a key person for sign off is away, the system ensures the request goes up to the next higher person for approval, rather than waiting on someone's desk for their return to the office. In Canada, if a request is for a data provider to receive a return of their own data, then the request is fulfilled for free. If it is a third party data request, then the first hour of work is free and any hours needed to prepare the approval or prepare the dataset are recovered at the rate of the analysts involved plus overheads (cost-recovery rates). The most complex requests have cost up to CAD 10-12 000.

In the Canadian province of Ontario, the Institute for Clinical Evaluative Sciences (ICES) is one of a small number of prescribed entities under Ontario law that may collect and process personal health information for research or statistical purposes. ICES launched a new data access service in 2014 to enable academic and non-profit researchers in Canada to apply for approval for access to de-identified linked record-level data within a secure data access environment. Since its launch there have been over 100 applications received. As ICES was established 22 years ago, it has developed metadata and data analysis programmes and tools to assist with analytical tasks. Tools that can assist external approved research applicants are made available to them. ICES has launched a pilot project with two commercial organisations to evaluate a possible extension of data access services to commercial applicants for approved research or statistical data uses. Results of the pilot are expected in 2015.

Under the funding provided to ICES from the federal and provincial governments to establish the data access service, applicants from Ontario academic and non-profit institutions can receive up to ten hours of a staff person’s time to assist with preparing their application for data access and preparing the data requested. After that time, applicants are charged for the staff time required to complete their request. In general, when requests are not complex, fees range from CAD 5 000 to 10 000. Fees for applicants from outside of Ontario are generally about 20% higher.

Countries with a small volume of data processing services

In Korea, most applicants for access to data are either HIRA staff or are researchers from other public authorities and from the academic/non-profit sector. Most applicants, 98% have been approved access. HIRA estimates that the elapsed time from a request to access to the data takes approximately three months. Efforts to expand the service and to recruit new personnel were put in place to reduce the elapsed time.

In Korea there are two staff members dedicated to processing applications for access to data and to preparing the data. In Korea, there is a fee charged for access to data and the average fee was estimated to be KRW 2million (about USD 1 380).

In Singapore, access to de-identified data for approved projects is only provided within the Ministry of Health’s secure microdata access lab. Generally there are no more than three researchers in the lab at one time. This is sufficient to meet the needs of researchers with approved projects and there is no waiting list. Researchers, whether ministry staff or external researchers with approved data linkage projects, must all do their analysis within the microdata access lab. Researchers with approved projects tend to be clinicians working within health care institutions, medical school researchers and university researchers. There are no fees charged for access to data. The ministry approves projects that it would support because it would derive a benefit from the study and will invest in facilitating the study.

In Switzerland, there is a legal ordinance that requires the FSO to send invoices for the time spent fulfilling applications for access to data. In practice, there is a lot of variability in how people estimate the working time spent on a data request. In general, the first half hour is free and so fairly simple requests, such as for access to data that are already public, are free. The cost of complex requests is determined on a case by case basis. There are also differences depending on the applicant. For example, students are charged a reduced fee and for-profit organisations are typically charged the full fee. In Switzerland, where data linkages tend to be probabilistic linkages, skilled statisticians are required to conduct the work and the ministry does not have very many experienced staff in this role.

Strategies and techniques to improve timeliness and reduce costs

Some countries have reported that resource constraints are a barrier to providing access to data and to supporting data processing requests. For example, Norway noted that there are various technical platforms in use across different data custodians that make data integration resource intensive and Spain indicated a shortage of resources to undertake data linkages, to de-identify data and, in particular, the lack of centralised authority to promote and facilitate secure data sharing and linkages as important obstacles to progress.

A few countries have reported new strategies that they have implemented to enable high quality linkages for approved projects at a lower cost and with a shorter time period for data processing. These initiatives are particularly useful in countries where key datasets have inconsistent identifiers or data quality problems with certain identifiers that result in more complicated and intensive statistical approaches to data matching.

Statistics Canada's Social Data Record Linkage Environment

Statistics Canada is Canada's national statistical organisation and it is the repository for personal datasets across various dimensions of social life which it collects and protects under the provisions of the *Statistics Act*. Statistics Canada recognises and supports record linkage projects but such projects have been very difficult and time consuming to conduct in the past. Statistics Canada is launching a new initiative to create a Social Data Record Linkage Environment. This is a secure environment in which an effort is made to develop a linkage key for data records within social datasets that could be included in approved data linkage projects in the future. A particular challenge in Canada is that there is no consistently captured person identification number among health datasets or among health and other social datasets. Instead, each dataset will have a number of differing person-identifiers, such as names, addresses, dates of birth, health insurance numbers, social insurance numbers and these will be captured with various levels of accuracy.

The census of population, taxation records, immigration records and birth and death records form the spine from which it is then possible for identifying variables within social datasets to be cleaned and a consistent unique identification key assigned. The unique identification key is stored separately from the original datasets and is available to be used only when a new data linkage project with that dataset is approved. A Privacy Impact Assessment has been prepared for the proposed environment and will require the approval of the Federal Privacy Commissioner.

With this proposed process, the datasets in the Social Data Record Linkage Environment are linkage-ready and the time and costs associated with record linkages are projected to be significantly lower after the environment becomes available. Prior to the creation of this environment, the costs of data linkage projects were very high and the annual number of approved applications was limited. The cost was typically in the range of CAD 150 000 to 200 000 with about 4-8 project applications submitted annually. After the creation of environment, the cost for a data linkage project is projected to decline to

CAD 20 000-50 000. Such a reduction will make it feasible for researchers with more limited project funding to place a data request. At the same time, Statistics Canada will have to adapt to accommodate a higher volume of requests, which is a concern given the limited number of staff with technical experience in conducting dataset linkages.

Automating data linkages: HSCIC England

At the Health and Social Care Information Centre in England, data linkages that take place regularly have been automated through the creation of computer algorithms that can be applied for approved projects. The automation of data linkages has enabled regularly occurring linkages to be done very efficiently and thus the processing times for data requests have improved. The cost of automation, in terms of time and resources, was incurred up front to develop and test the computer code. The code links the data and then provides an output indicating how many records linked perfectly and how many were linked with a less than perfect match/lower level of reliability. The computer code creates a variable within the dataset that indicates the strength of the linkage for each record so that when the data are analysed, researchers have the option to restrict their analysis to only the strongest links. The computer code checks for known data quality problems, such as not permitting a link to a death record for an individual who appears at a later time still alive and in hospital. Automation contributes to improvement in the underlying data quality. For example, when the data warehouse for HES was created, automated reports about data quality were produced leading to data providers improving the underlying data submitted. More could still be done with the data linkage automated reports, in terms of communicating data quality problems back to data providers.

Accreditation or certification of data processors promotes both data security and access to data

Concentration can have positive implications for the protection of health information privacy. Firstly, an accreditation or certification process can narrow the number of processors to only those who meet the country's highest standards for data privacy and security protection. Secondly, accredited organisations can be resourced to develop the highest levels of data privacy protection including sophisticated data de-identification processes and secure access mechanisms such as secure real-time remote data access and supervised research data centres, as will be discussed in Chapter 8. Thirdly, the designated organisation can ensure that all staff has the appropriate training and skills to meet the data confidentiality, privacy and security requirements of accreditation. If not, the organisation loses its accreditation.

Concentration can be argued as being a risk to data privacy and confidentiality protection, as a data breach from a large processer can be more damaging than a breach from a small processer with few data records and a larger processor has greater capacity to re-identify previously de-identified data. However, with good governance mechanisms assured through accreditation or certification, the risks from concentration can be managed. Much can be learned from national statistical authorities who regularly process and protect each country's sensitive population-level personal data.

Countries that have concentrated national health data custodianship and processing have a strong technical advantage in linking data across the continuum of health and health care experiences in their populations and are much better able to develop health care quality and performance monitoring indicators and research programmes. It is much easier for custodians of concentrated data to initiate data linkage projects because it is not necessary to first open a negotiation for data sharing among several organisations to enable work to be undertaken. Further, custodians of concentrated data can be appropriately resourced to develop the human resources with the skills necessary to efficiently and accurately process the data and can build unique knowledge about the quality, limitations and statistical and research possibilities that lie within the nation's health data resources. Thus they are better able to support secure, timely and affordable access to data for approved data requestors.

Countries have invested heavily in the development of national datasets and in electronic health record systems and it makes economic sense for the data from these systems to be used to its fullest potential to benefit societies, provided that there are national data governance frameworks that safeguard the privacy of data subjects. Accommodating requests for access to data from external applicants will require resources within data custodians and processors. The onus is on custodians and processors to make their data processing services as efficient as possible and the onus is on governments to assure that custodians and processors are adequately resourced to set up efficient processing systems and that access to data will be fairly provided to applicants with approved projects, including key health system stakeholders. In so doing, research and statistics for the benefit of society can develop. The remainder of this document describes elements of the strong governance of personal health data that should be an integral part of countries' requirements of the custodians and processors of national personal health data.

The Advisory Panel of Experts on Health Information Infrastructure identified accreditation or certification as mechanisms supporting privacy-protective data use and identified the following key features of accreditation or certification:

4. A certification/accreditation process for the processing of health data for research and statistics is implemented

The certification/accreditation process

a) Limits processing of identifiable data and data linkages to certified/accredited data custodians and processors.

b) Requires certified/accredited data custodians and processors to comply with norms for data governance that, such as the eight data governance mechanisms identified in this report.

c) Establishes rules, policies, data standards and administrative structures among certified/accredited data custodians and processors that encourage and support appropriate co-operation for data sharing and analysis that minimise barriers.

d) Requires certified/accredited data custodians and processors to act as a secure national archive for personal health data with future research and statistical value.

e) Adequately resources and requires data custodians and processors to ensure that any fees to process data requests do not limit fair access to data for approved applicants from all sectors of society.

f) Requires accountability for adherence to certification/accreditation norms and for the timeliness and quality of data processing services.

References

Australia (2014a), "Statistical Data Integration Involving Commonwealth Data", www.nss.gov.au/nss/home.nsf/NSS/0E887A88A9224F8BCA2577F20016FE5D?opendocument, accessed 30 July 2014.

Australia (2014b), "Accredited Integration Authorities: Australian Institute of Health and Welfare Application Summary", http://nss.gov.au/nss/home.nsf/pages/Data%20Integration%20-%20AIHW%20accreditation%20application%20and%20audit%20summary/, accessed 30 July 2014.

IGRP – Information Governance Review Panel (2013), "Dame Caldicott Chair Information to Share or Not to Share: The Information Governance Review", www.gov.uk/government/uploads/system/uploads/attachment_data/file/192572/2900774_InfoGovernance_accv2.pdf, accessed 1 August 2014.

SAIL – Secure Annonymised Information Linkage (2014), "Secure Annonymised Information Linkage Databank", Wales, www.saildatabank.com, accessed 4 August 2014.

SHIP – Scottish Health Informatics Programme (2014), "Scottish Health Informatics Programme", www.scot-ship-toolkit.org.uk/route-maps/researchers-data-security/route-a/data-access-though-safe-haven#what-is-a-ship-safe-haven accessed 4 August 2014.

SHIP (2010), "SHIP Guiding Principles and Best Practices", Scottish Health Informatics Programme, www.scot-ship.ac.uk/sites/default/files/Reports/Guiding_Principles_and_Best_Practices_221010.pdf, accessed 4 August 2014.

Chapter 6

Fair and transparent health project approval processes

This chapter describes the project approval processes in place in the OECD countries participating in this study and the public transparency of those processes.

A Risk-Benefit Evaluation Tool is presented that countries could use to guide bodies evaluating and approving applications to process personal health data. The chapter emphasises the importance of the consideration of both societal risks and societal benefits when decisions are taken.

Highlights

Fair and transparent project approval processes are essential to meeting public expectations regarding appropriate uses of their personal health data. Such objectives are more attainable when decision making is supported by an independent, multidisciplinary project review body such as a research ethics committee.

Elements of a fair and transparent process include the independence of project reviewers from those seeking or realising a benefit from a project; the degree to which the public is informed about the existence of, the members of and the role of project reviewers; and the degree to which national authorities are open about the process that must be followed to apply for and be approved access to de-identified data, including data from record linkage processes.

Among the countries participating in this study, there were five different approaches to project approval decision making described by experts: independent research ethics review boards that advise data custodians on proposals involving the processing of personal health data; internal committees or governing boards of data custodians that take decisions on project approval and have a mix of both internal and external experts; data protection regulators who take the final decision on project approval with advice from research ethics boards and data custodians; data custodians who take the final decision after consultation with the data privacy regulator; and data custodians who take the final decision after an internal process.

The degree to which information is provided to the public about the process to request access to de-identified data or to request a dataset linkage is variable among the countries. Nine countries indicated that there is public information, such as a website where the process for requesting access to de-identified data and the process to request a dataset linkage are described for all of the key national health datasets [Czech Republic, Iceland, Korea, New Zealand, Norway, Sweden, United States, United Kingdom (England and Scotland)] or for the majority (Canada, Netherlands, and Finland). Five countries are fully transparent about applications for access to de-identified data but, conversely, do not provide information about the record linkage process for all or most key health datasets (Denmark, Ireland, Switzerland, United Kingdom (Wales), and Japan). Five countries provide little to no information (Italy, Singapore, Israel, Spain and Turkey).

As part of this OECD study, the members of the Advisory Panel of Experts on Health Information Infrastructure developed a Risk-Benefit Evaluation Tool that could be used as a guide by bodies evaluating and approving applications to process personal health data (Table 6.2). The tool identifies areas of both societal benefit and societal risk that should be considered as part of the decision-making process. Decision making will be unbalanced and, therefore, likely sub-optimal for the society, when decisions consider only one of these two dimensions.

Project approval processes that are fair and transparent and that take ethical considerations as well as legal requirements into account are essential to meeting public expectations regarding appropriate uses of their personal health data. This chapter describes the project approval processes of the OECD countries participating in this study including the involvement of research ethics boards, internal review boards and data privacy regulators in decision making about the approval of projects involving the processing of personal health data; and public transparency regarding the project approval process. A *Risk-Benefit Evaluation Tool* is presented that countries could use to guide bodies evaluating and approving applications to process personal health data (Table 6.2). It emphasises the importance of the consideration of both societal risks and societal benefits when decisions are taken.

Project approval processes

Among the countries participating in this study, there were five different approaches to project approval decision making described. There are examples where independent research ethics review boards advise data custodians on proposals involving the processing of personal health data (Israel, Sweden, New Zealand, Norway, United Kingdom and the

United States). There are also examples where data custodians have established internal committees or governing boards that take decisions on project approval and have a mix of both internal and external experts (Japan, Korea and the United States). There are examples where the data protection regulator takes the final decision on project approval with advise from research ethics boards and custodians (Denmark and Iceland) and where the regular is consulted and provides advice to the data custodian (Finland, Switzerland, Czech Republic). There are also examples where decisions are typically taken through the data custodians' internal processes (Canada, Netherlands, Spain and Singapore).

Research ethics committees

There are international ethical guidelines for biomedical research involving human subjects published by the Council for International Organizations of Medical Sciences and the World Health Organization (CIOMS, 2002). These guidelines provide an international consensus view of the role of research ethics boards. They state that "all proposals to conduct research involving human subjects must be submitted for review of their scientific merit and ethical acceptability to one or more scientific review and ethical review committees. The review committees must be independent of the research team, and any direct financial or other material benefit they may derive from the research should not be contingent on the outcome of their review. The investigator must obtain their approval or clearance before undertaking the research. The ethical review committee should conduct further reviews as necessary in the course of the research, including monitoring of the progress of the study." The role of research ethics boards is described as inclusive of granting a waiver for exemption to patient consent requirements, including cases of the secondary use of patient data.

The same groups have also developed international ethical guidelines for epidemiological studies (CIOMS, 2008). These guidelines reinforce the 2002 guidelines by clarifying that epidemiological studies require research ethics and scientific review and that review committees should hold the role of granting a waiver for exemption to consent requirements, unless such a waiver is already granted by a national legislation that is following the same ethical principles as are laid out in their guidance.

A number of countries have established research ethics review boards (REB) that play a role in decision making about projects involving the processing of personal health data, such as data linkage studies. Often the REBs have other roles as well, such as approval of the creation of new national datasets, approval of uses of biological samples or approval of clinical intervention research projects.

In Sweden, if a proposed research project involves the personal health data of living persons then the application to conduct the project must first be approved by a research ethics board. In Sweden, there is a regional research ethics board in each region of the country. Researchers conducting a project in a single region apply to the regional REB. In cases where the researcher's project is national or multi-regional, each REB would be asked to approve the project. Once a project has received the required REB approvals, then the next step is to apply to the data holders for their approval. Researchers sometimes believe that once an REB approval is granted their project is approved and are surprised that the data holder's approval is also needed and can differ from the REB decision. The same approval process applies to all personal health data.

In Sweden, approval is more likely to be granted if the request is for de-identified data. There is a check made to ensure that the request to the National Board of Health and Welfare (NBHW) matches the application for REB approval. Both the REB application and the REB decision are reviewed by the NBHW. In some cases, researchers ask the REB for access to an entire database. The NBHW would be unlikely to approve this request, even if there is an REB approval. The NBHW will ask the applicant to narrow their request to fit the purpose of their study, such as selecting patients with certain diseases and identifying the variables they need for their analysis to be conducted. The NBHW consults with the Data Protection Authority for advice, as needed.

In Israel, there are internal research ethics review boards within each hospital and there is also one National Research Ethics Review Board. The National REB hears all requests for access to genetic data, studies involving assistive reproductive technologies, all research studies undertaking by the Ministry of Health and any project considered to be new or as having a potentially wide impact. Before the National REB hears an application, it must first be approved by the internal ethical review boards within each hospital whose data are involved in the project. If a study is a more typical study undertaken by a hospital, then it may be only heard by the hospital's internal research ethics review board. Israel notes that the research ethics boards consider whether or not the data use should be permitted, whether or not patient consent is required, whether or not identifiable data are needed, and how data will be de-identified.

In New Zealand, if a researcher is requesting access to identifiable personal health data, then they must have the approval of a research ethics committee that is independent from the researcher and the research project. The national research ethics committee would review proposals for projects involving more than one health region. Research within a single region would be proposed to the ethics committee for that region. As a result, there is only one ethical review needed for each project proposed. Research ethics committees can be established by universities, other government departments and private organisations. There is wariness of the independence of ethics committees established by the private sector and, particularly, when members are paid.

In England, there are 3-4 different research ethics boards (REBs), all of which are external to the Health and Social Care Information Centre. Each REB can provide the research ethics approval necessary before the HSCIC could approve access to data. REBs evaluate the purpose and methods of proposed projects and determine if the benefit of the project outweighs the impact of the project on data subject's privacy. REB approval is necessary before the HSCIC can approve access to identifiable data and to de-identified record level data that has a high re-identification risk. Following the introduction of the *Care Act* in 2014, the Confidentiality Advisory Group of the Health Research Authority has a statutory mandate and the mandate includes advising the HSCIC on the disclosure of data.

In the UK Wales SAIL project, applications for access to data are reviewed by an Information Governance Review Panel which is a body independent of SAIL. Members of the panel include the custodians of the datasets within SAIL including the NHS Wales Informatics Service, the British Medical Association, the National Research Ethics Service and members of the SAIL consumer panel. The Information Governance Review Panel advises on whether the proposed project is of sufficient public interest to be approved; how the proposal may be modified to reduce re-identification risk; and whether the project needs a research ethics approval before it can be approved. As the SAIL system only contains de-identified data, requests for data linkages and access to SAIL data do not require the approval of a research ethics body, because they do not involve the processing of identifiable data. If the researcher has a cohort of data they have collected that they wish to have linked to SAIL data; however, they must also provide documentation of research ethics approval and patient consent.

Research ethics approval is a requirement for the approval to process data without consent in the United States. The environment for obtaining research ethics approval is fragmented. Research requiring health data from different states and providers can require the approval of all of the research ethics boards at the state level as well as research ethics boards within the different health care providers whose data would be involved in the project. Each IRB sets different conditions for approval and has different processes and application forms, so the process of obtaining approval is complex and slow. In the United States, the states control the uses of their populations' data and the state laws differ in terms of certain sensitive topics in health that they may not approve, such as research into abortion. Further, health care providers control their own data and may not agree to even a state-level IRB deciding on projects on their behalf. Centralising to one IRB or even one IRB per state is legally possible but politically hard. Nonetheless all of these IRBs are following the common rule.

In Norway, there are seven Regional Committees for Medical and Health Research Ethics (REC) that are founded upon the Norwegian *Law on Research Ethics and Medical Research*. These RECs are made up of people with different professional backgrounds including lay representatives and representatives of patient groups. The RECs advise data custodians on proposals involving the processing of personal health data.

At the Institute for Clinical Evaluative Sciences in the Canadian province of Ontario, external applicants for access to de-identified person-level data must secure approval for a proposed project from a research ethics board before ICES staff will approve their application for access to data.

Approval by the Data Protection Regulator following input from research ethics boards

There are two examples where the data protection regulator takes the final decision on project approval.

In Iceland, the Data Protection Authority would approve the application, conditional upon confirmation that the data custodians involved approve the project and in consideration of the opinion of the National Bioethics Committee. All scientific research involving personal health data needs to be approved by the National Bioethics Committee, the data custodians involved and the Data Protection Authority.

In Denmark, the Data Protection Authority (DPA) approves requests for 1) linkages of a registry to another registry or dataset and 2) the creation of a new registry. The DPA is responsible for the criteria it uses to evaluate requests. The project proposed must have a clear purpose and not be too broad. The data requested must be limited to what is needed to fulfil the purpose of the project (using as little data as is possible). There are two national research ethics boards, one for health data and another for other data. Some registries and databanks require applicants to seek a research ethics board approval before an approval from the DPA can be granted. The SSI can approve projects involving access to its individual registries without seeking DPA approval. There are no differences in the approval process by the type of data requested. For example, the SSI maintains a national bio bank and, at present, applications for the linkage of bio bank data to other health registries proceeds in the same manner as applications not involving bio bank data.

Independent advisors within internal committees or governing boards

There are examples where data custodians have established an internal project review committee and invited independent experts to take part in the committee.

In Japan, the Insurance Bureau reviews applications for access to data for their methodology and awareness of the strengths and limitations of the insurance claims data. Applications that are reasonable are then reviewed by a governing board that makes the final decision on applications for access to data. The Bureau's governing board includes representatives from the medical, pharmacy, and dentistry associations and researchers. The board reviews applications for access to data twice per year, with about 10-15 applications reviewed at each meeting. There are two key criteria for approval: 1) content of the application and 2) compliance with data security requirements. For content, the application must describe the project and the project methodology must be feasible to conduct with insurance claims data. The claims data are difficult to work with and because the data have only recently been made available to researchers, statistical literacy regarding these data is still developing. The Bureau works with applicants to define projects that could be feasible. For example, there was an application to use the data to evaluate the effectiveness of smoking cessation treatments before and after the treatments were introduced. However, as there is only five years of patient history in the database, there is not enough follow-up time to evaluate the effectiveness of the programme yet. Universities must meet data security requirements for their researchers to gain access to data. The Bureau conducts security audits of the universities to ensure that the secure room and procedures conform to Bureau requirements.

In the United States, the NCHS has an internal review board called the Ethics Review Board or ERB. The ERB is composed of senior officials of the NCHS and external experts. Every five years, the NCHS obtains approval from the ERB for its data linkage programme and data access practices. Under this ERB approval, key NCHS officials undertake the review and approval of applications for access to data. This includes the managers of any of the datasets involved in the request, the NCHS confidentiality officer and the director responsible for data access. Each reviews the request. If the request is of a routine nature, and is for a statistical purpose, it would be approved. If the request raises concerns of any of these officials, then the request is either referred to the ERB for approval or the requestor is asked to revise their application and resubmit it. The ERB meets monthly and a large share of its activities are related to requests for access to genetic data and biological samples. Applications for access to data are not judged on their scientific merit. If data from the NCHS is to be linked with data from other public authorities, the legal requirements of all of the data custodians involved must be respected.

In Korea, for requests to access the data holdings of HIRA, there are two approval committees. The first committee consists of managers in the IT section, the section on personal information management and the section on statistics. The second committee is composed of 50% HIRA managers, and 50% experts from outside of HIRA including the Medical Association, the Pharmaceutical Association, and civic groups. This second committee is convened on a case by case basis to advise on new or unusual requests for data processing and access to data. HIRA can approve requests to link its data holdings with data collected by a researcher, such as clinical trial data. In cases where the data of HIRA are requested to be linked to data from another sector, the committee(s) of HIRA would need to approve the application along with the relevant committees in the other organisations whose data would be involved.

Internal decision-making process with advice from the privacy regulator

In Finland, under the *Personal Data Act*, anyone requesting access to personal health data or a data linkage of personal health data needs to submit an application that includes a research plan. For THL, the application is reviewed by the Data Protection Ombudsman who provides an opinion based on law to THL. The Ombudsman considers, for example, whether it would be necessary to provide identifiable data for the research use described. Then all of the data holders involved have the responsibility of making the final decision on whether or not the data use will be approved. The data holders use their experience to evaluate whether the research purposes described would be appropriate. A review by a research ethics committee is not required when the request is for access to data from existing registers or administrative datasets. A research ethics review is required if the project will involve genetic data or a clinical intervention. If data are to be collected directly from individuals as part of the project, then they must be informed about the data linkages that will be conducted with the data.

In Switzerland, the FSO has the sole responsibility to evaluate and approve applications for access to data. The FSO has an internal committee involving different sectors and legal experts. The committee is available to provide advice to managers and provides managers with non-binding recommendations. Depending on the project, the FSO sometimes consults with the Data Protection Authority or with other ministries, particularly the Federal Office of Public Health. In practice, the managers of each dataset make the approval decision in consultation with their section head (supervisor). If a decision is not clear, then the supervisor will seek advice or seek approval at a higher level in the organisation. In cases where a request would involve the data holdings of more than one ministry, such as health and education data, then the approval decision would be made at the Director level in both organisations.

In the Czech Republic, the ministry makes the final decision about approving projects involving the processing of personal health data and is responsible for approving projects within the requirements of the laws. The Data Protection Office can be asked to provide advice on complex cases. Also, for every registry there is an advisory board that includes representatives from the Ministry of Health and the Medical Society. Similar to a research ethics board, these advisory boards provide recommendations on how to handle requests. The advisory board for the cancer registry, in particular, frequently discusses requests for data from foreign entities.

Internal decision-making processes

In Canada, if the request is from a data provider organisation to CIHI and that organisation is requesting a return of their own data, then the request would be routinely approved by a director. If the request is from a third party, such as an academic researcher, then there is a more stringent approval process. The applicant must explain the purpose of the research and must justify the datasets and the variables within the datasets that are required to fulfil the research. A data inquiry form is provided on the CIHI website that the researcher uses to make a data request. The researcher will then enter into a discussion with a CIHI analyst to ensure that the data requested matches and does not exceed what would be needed to undertake the project. The request is then reviewed by an internal CIHI committee.

In the Netherlands, the data owners involved in the request must all approve access to their data. For example, for a data linkage project involving Dutch hospital data, the data owner must approve and then Statistics Netherlands, who holds the Dutch hospital data, must also approve. There are also research ethics committees in the Netherlands involved in approving studies involving clinical interventions. Studies involving the linkage of existing datasets do not seek research ethics approval.

In Spain, each data custodian is responsible for granting approval for access to data. The Ministry of Health has established an internal committee that reviews requests for access to de-identified microdata from its health care registries.

In Singapore, the data custodian makes the decision on data disclosure. Depending on the project there may be consultation with other groups. For example, for a clinical improvement agenda there was consultation between clinicians and data custodians regarding data sharing and access.

Appeals process

A few countries explained the processes that could be followed by an applicant who is not satisfied with a project approval decision.

In Israel, if an applicant is not satisfied with a project approval decision taken by the public sector, the applicant can challenge the decision in court. If the decision was taken by a private organisation, there is no legal recourse for the applicant. A data requestor who has been denied access to Canadian data from CIHI may apply to individual provinces for their approval and access data via them. In Korea, the applicant would appeal to the HIRA committee. If they were unsatisfied with the response, they would have the option to appeal to the government. In Finland, once a decision has been taken, the requestor has 30 days to appeal to the court if they are not satisfied with the decision. In the United States, if a researcher disputes with an NCHS manager regarding their application for access to data, the Internal Review Board will render a decision on the application and its decision is binding. There is a National Research Ethics Board (REB) in Sweden and the role of this REB is to hear appeals from researchers whose projects have been refused by a regional REB. In Sweden, when the NBHW has rejected a request for access to data, it is possible for researchers to appeal to a court.

Transparent processes for requests to process or access personal health data

The degree to which information is provided to the public about the process to request access to de-identified data or to request a dataset linkage is variable among the countries participating in this study. Eight countries indicated that there is public information, such as a website, where the process for requesting access to de-identified data and the process to request a dataset linkage are described for all of the key national health datasets (Table 6.1). Four countries make such information available for the majority of their national datasets. Five countries are fully transparent about applications for access to de-identified data but, conversely, do not provide information about the record linkage process for most or all of the key health datasets. Five countries provide information for only a minority or none of the key national datasets.

Table 6.1. Public communication regarding requests for access to and processing of personal health data

	Proportion of key national personal health datasets with a…	
	Published process for requesting access to de-identified data	Published process to request a record linkage
Czech Republic	100%	100%
Iceland	100%	100%
Korea	100%	100%
New Zealand	100%	100%
Norway	100%	100%
United States	100%	100%
UK England	100%	100%
UK Scotland	100%	100%
Sweden	89%	89%
Canada	88%	88%
Netherlands	86%	71%
Finland	78%	78%
Denmark	78%	11%
Ireland	80%	20%
Switzerland	100%	0%
UK Wales	100%	0%
Japan	86%	0%
Italy	29%	0%
Singapore	25%	25%
Israel	0%	0%
Spain	0%	0%
Turkey	0%	0%

Source: Author’s own calculations based on the results of this study.

Project review boards must evaluate the risks and benefits to society of a proposed use of personal health data

As part of this OECD study, the members of the Advisory Panel of Experts on Health information Infrastructure developed a *Risk-Benefit Evaluation Tool* that could be used as a guide by bodies evaluating and approving applications to process personal health data (Table 6.2). The tool identifies areas of both societal benefit and societal risk that should be considered as part of the decision-making process. Decision making will be unbalanced, and therefore sub-optimal and risky, whenever decisions consider only one dimension. As was discussed in Chapter 1, when decision making is unbalanced in the extreme, decisions can be taken that either result in such limited data that the health of populations and the quality and efficiency of health care are harmed or such liberal access to and use of data that individual’s privacy is violated with subsequent harms to people and loss of public confidence.

Table 6.2. Risk-benefit evaluation tool supporting decision making about the processing of personal health data

	Societal benefits		Societal risks
1)	Is the data use a/an:	7)	What is the Identifiability of the data required to successfully undertake the project?
a)	Ad hoc/one-time only research or statistical project?	a)	Aggregated data that could be made public (anonymised data)
b)	Part of an on-going programme of scientific research?	b)	Anonymised micro data treated to protect against re-identification that could be made public (public-use micro data)
c)	Part of regular reporting of statistics or indicators for monitoring?	c)	De-identified micro data where ID numbers and other direct identifiers are encrypted or suppressed, and potentially identifying variables have been treated (aggregation, masking, swapping, suppression)
d)	To create or enhance an on-going data-set or registry?	d)	De-identified micro data where ID numbers and other direct identifiers are encrypted or supressed
2)	Is the data use consistent with acceptable uses of the data?	e)	Micro data with identifiers included (fully identifiable data)
3)	What will be the potential benefits of the project? Will results improve:	8)	Could the objectives of the study be realized if at any stage of the project, individual data are aggregated, stored and exchanged in aggregated format only?
a)	Health outcomes?	9)	Could a sample be drawn from the data or is full population data necessary?
b)	Treatments?	10)	Have data subjects consented to the processing?
c)	Patient health care experiences?	11)	Is the collection of informed consent of data subjects practicable to successfully undertake the project?
d)	Quality of health care?	12)	Is an exemption to patient consent requirements legally permissible?
e)	Efficiency, cost or affordability of health care?	13)	Are all elements necessary to grant an exemption to patient consent requirements fulfilled?
f)	Management or governance of the health sector?	14)	Is it necessary to seek the advice of a research ethics board or committee?
g)	Profits or market share for individual health system actors?	15)	Has a research ethics board rendered a positive decision?
h)	Growth of the health care industry or the economy?	16)	Is it necessary to seek the advice or decision of a data protection authority?
i)	Progress of science, research, or innovation?	17)	Has the data protection authority rendered a positive decision?
j)	Quality of health statistics?	18)	Have the custodians of the data involved rendered a positive decision?
k)	Expense or respondent burden of alternative data collection methods?	19)	Has a risk analysis (meeting appropriate standards) been done?
l)	Transparency or accountability of government programmes?	20)	Does the applicant have a track record of privacy protective data use?
4)	Who are the potential beneficiaries of the project results? Are they	21)	Would the data recipient fall under any legal requirements to protect the privacy of data subjects?
a)	Multiple societies/global population?	22)	Are there legal sanctions that could be applied if the data was misused by the requestor?
b)	Society/whole population?	23)	If a foreign applicant, does the legislative framework for the protection of data privacy in the foreign country adequately meet the legal standard of the home country?
c)	Patient groups?	24)	Is it necessary to transfer the data requested to the data recipient?
d)	Government/policy makers?	25)	Could a research data centre or secure remote data access system be used to provide the recipient with access to the data?
e)	Research community?	26)	If it is necessary to transfer the data….
f)	Health care industry?	a)	How will the data be protected during the transfer process?
5)	What may be the potential impact of the project results on beneficiaries?	b)	Are the data requestor's physical security and security policies and practices sufficient to mitigate risks?
6)	Are the proposed data sources and methods appropriate to realise the potential benefits?	27)	How vulnerable is the data to an outside attack during the transfer process?
		28)	How vulnerable is the data to an outside attack on the data security environment of the data requestor?
		29)	If there was a successful attack from the outside, how difficult or expensive would it be for the hacker to identify or re-identify data subjects?
		30)	What could be the harms incurred if an outside attack were successful?
		31)	How long will identifiable data (or data with a high re-identification risk) be kept before it is either anonymised or destroyed?
		32)	If approved, what will be the process used to follow-up with the data requestor to ensure that all of their legal and contractual obligations have been respected?

Source: Author's own work.

Transparent and fair project approval processes are needed

Fair and transparent project approval processes are essential to meeting public expectations regarding appropriate uses of their personal health data. Elements of a fair and transparent process include the independence of project reviewers from those seeking or realising a benefit from a project; the degree to which the public is informed about the existence of, the members of and role of project reviewers; and the degree to which national authorities are open about the process that must be followed to apply for and be approved access to de-identified data, including data from record linkage processes.

The Advisory Panel of Experts on Health Information Infrastructure identified the following key features of the governance of health data that promote openness and transparency:

5. The project approval process is fair and transparent

The project approval process

a) Follows a criteria for project approval that considers both societal risks AND societal benefits of proposed data uses, such as the risk-benefit evaluation tool included in this report.

b) Considers the elements of the proposed statistical or research use of data on their own merits and avoid discrimination against applicants due to their age, experience, employment or other factors.

c) Ensures the process to apply for approval to process and/or access personal health data and the criteria for project approval are publicly available (such as a website).

d) Ensures a summary of each application for project approval and each approval decision are publicly available (such as a website). Summaries include the purpose of the processing, the datasets included and the organisations and researchers involved.

5.1. A multidisciplinary project approval body

a) Includes relevant stakeholders, such as legal experts, privacy experts, statistical experts, patients and researchers that are also third parties, with no stake in an approval decision.

b) Consults with the custodians of all datasets involved in a proposed project and takes their advice into account.

c) Is publicly identified, including the project approval body's role, membership, criteria the body follows for project approval, timeliness of approval decisions, and process to appeal a decision.

d) Is accountable for the timeliness and quality of their services.

References

CIOMS – Council for International Organizations of Medical Sciences and WHO (2008), *International Ethical Guidelines for Epidemiological Studies*, Council for International Organizations of Medical Sciences in collaboration with the World Health Organization, Geneva.

CIOMS and WHO (2002), *International Ethical Guidelines for Biomedical Research Involving Human Subjects,* Council for International Organizations of Medical Sciences in collaboration with the World Health Organization, Geneva.

Chapter 7

De-identifying personal health data

This chapter discusses examples of how data de-identification is applied in practice among custodians of national health datasets in the OECD.

The chapter notes gaps between the goals of legislative requirements for data protection and decision making about data de-identification methods in practice. It describes country reports of de-identifying national datasets prior to analysis and describes the use of pseudonymisation to replace direct patient identifiers in national datasets. It discusses how countries apply other data de-identification techniques to address the risk of data becoming re-identified and reviews the importance of weighing data de-identification techniques against the utility of the dataset for its intended purpose.

Highlights

National personal health datasets are often identifiable, either directly, because they contain information identifying data subjects such as names, addresses and ID numbers, or indirectly, because variables within the dataset can be used to infer the identity of individuals. Data are considered to be de-identified when they do not identify individuals directly and they cannot reasonably be used to determine individuals' identities.

Countries have introduced data de-identification practices in order to make it more difficult to identify individuals within a dataset. Rarely, however, will data de-identification processes reduce to zero the risk that an individual could be identified from the data. This is because the data can still be sufficiently detailed for it to become re-identified by someone with knowledge about the data subjects or because it is linked or merged with other identifiable data.

There is often a gap between the letter of the law with respect to the protection of privacy and the application of data de-identification methods to protect the identities of data subjects. This is because rarely do legal frameworks provide detailed guidance regarding when to conclude that de-identification has rendered data anonymous and when it has left the data too risky to be outside of legal protection.

In many countries, identifying information about individuals, such as names, complete addresses and ID numbers are converted to a meaningless name or number in a consistent manner that permits record linkage among databases for approved projects.

Another technique is "data masking", which involves modifying a wide range of dataset variables in order to reduce the likelihood that these variables could be used to re-identify the data. Suppression is a common data masking technique that describes the complete removal of information from a dataset.

Decisions about data de-identification need to consider "the big picture" of data protection, security and utility. Data suppression and masking techniques can have detrimental impacts on the ability to conduct certain studies or on the validity of study findings when the data are analysed. Data security techniques discussed in the next chapter can be used to protect against data re-identification.

Data de-identification methods that are satisfactory today will need to be revised with the introduction of new technologies, new health data and new data privacy protection risks.

National personal health datasets are often identifiable, either directly, because they contain information identifying data subjects, or indirectly, because variables within the dataset can be used to infer the identity of individuals. Countries have introduced data de-identification practices in order to make it more difficult to identify individuals within a dataset. Rarely, however, will data de-identification processes reduce to zero the risk that an individual could be identified from the data. This is because the data can still be sufficiently detailed for it to become re-identified.

As a result, data de-identification processes are often applied in combination with other data governance and security measures to create a secure environment within which personal health data may be accessed and analysed for approved purposes (see Chapter 8). When the environment within which analysis is permitted has been made secure, the degree of data de-identification steps required can be lessened to enable the required data use.

This chapter discusses examples of how the broader security environment governing the data should influence decision making about data de-identification methods; notes gaps between the goals of legislative requirements for data protection and decision making about data de-identification methods in practice; describes country reports of de-identifying national datasets prior to analysis; describes the use of pseudonymisation to replace direct patient identifiers in national datasets; discusses how countries apply other data de-identification techniques to address the risk of data becoming re-identified; and reviews the importance of weighing data de-identification techniques against the utility of the dataset for its intended purpose.

Gap between legal requirements and data de-identification in practice

In most cases, there is an important gap between data privacy protection legislations and the application of data de-identification techniques. This is because rarely do legal frameworks provide detailed guidance regarding how to approach data de-identification and when to conclude that such de-identification has rendered data anonymous and when to conclude that the data should still be considered personal and under the protection of the data privacy law. Exceptionally, the main national health information privacy law in the United States (HIPAA) is specific about what it means to de-identify personal health data. Section 164.514(a) of the HIPAA Privacy Rule describes the standards for data de-identification and includes two different methods (DHHS, 2012):

Method 1: This method is referred to as the safe harbour. There are 18 specific data fields that must be suppressed for personal health data to be defined as fully de-identified data under the HIPAA privacy rule. These fields include direct identifiers: names, addresses, postal code, telephone numbers, fax numbers, social security numbers, medical record identifying numbers, biometric identifiers, photographs etc. Also in the list are full calendar dates and single years of age for persons after age 89. Researchers with the need for access to exact dates or postal codes must apply for research ethics board approval as such data are considered identifiable.

Method 2: An expert determines whether the risk of data re-identification is very small. Method 2 may allow for less de-identification and therefore under this method there may be higher re-identification risk. This is because experts in data de-identification have different views about methods to be used and about the re-identification risk remaining after different methods are applied to data.

Even with the specificity included within HIPAA method 1, the risk of dataset re-identification has not been eliminated. Re-identification is attributing identifying variables to an individual's record within a de-identified dataset. Re-identification requires information about the individual obtained from personal knowledge or from data stored in other datasets about the same individual. For example, a person who is listed in a non-health dataset with their name and address included might be matched, with some probability, to a health dataset that has no names or addresses included. Using probabilistic record linkage, the two databases are linked to the same individual on the basis of similar variables available in both datasets. Examples of similar variables might be city, sex, age, marital status, diagnosis, etc.

Effective governance of personal health data requires awareness and evaluation of privacy risks that are inclusive of both data de-identification and the broader data governance and security environment protecting the data.

Five key de-identification practices were explored in the OECD country survey (Table 7.1) and each will be discussed in detail in this chapter.

Table 7.1. Proportion of key national health datasets with five data de-identification practices

	Data are de-identified prior to analysis	Data de-identification method is documented	Pseudonyms are created from direct identifiers	Risk of data re-identification is assessed	Usefulness of the data for the planned analysis explicitly considered
Singapore	100%	100%	100%	100%	100%
Czech Republic	100%	100%	100%	100%	100%
UK Scotland	50%	100%	100%	100%	100%
UK Wales	50%	100%	100%	100%	100%
Italy	100%	100%	29%	100%	100%
United States	100%	100%	14%	100%	100%
Korea	100%	100%	100%	0%	100%
Norway	100%	50%	80%	100%	100%
Netherlands	100%	57%	71%	71%	71%
New Zealand	50%	0%	100%	100%	100%
Sweden	44%[1]	0%	89%	89%	89%
Japan	86%	86%	86%	86%	0%
Spain	75%	75%	0%	75%	75%
Finland	39%	0%	78%	78%	78%
Canada	88%	88%	0%	88%	0%
Iceland	0%	80%	60%	40%	80%
UK England	40%	60%	60%	60%	20%
Denmark	67%	67%	67%	11%	0%
Turkey	0%	80%	0%	80%	0%
Switzerland	20%	20%	0%	0%	20%
Ireland	50%	0%	0%	0%	0%
Israel	0%	0%	0%	0%	0%

[1] Datasets are usually, but not always, de-identified prior to analysis.

Source: Author's own calculations based on the results of this study.

Data are de-identified prior to analysis?

Countries were asked whether or not key health datasets are de-identified prior to analysis (Table 7.1). Seven countries responded that for all of their key national health datasets, the data are de-identified prior to analysis (Czech Republic, Italy, Korea, Netherlands, Norway, Singapore and United States). On the other end of the spectrum, four countries reported that data is not de-identified prior to analysis for any of the key national health datasets (Iceland, Turkey and Israel). In general, countries that provide mechanisms for third parties to be approved access to de-identified data have developed awareness of and practices to protect data while it is being analysed, whether analysis takes place within the data custodian or in an external organisation. Some countries reported that data are usually but not always de-identified prior to analysis because requests for access to identifiable health data for statistics or research purposes may be approved (Finland, New Zealand, Sweden and the United Kingdom). Other countries were not certain of data de-identification procedures for all of their key national health datasets.

Countries applying de-identification methods tended to also report that they have documented at least some aspects of these methods and a few shared documents that are available in English (CIHI, 2014; HSCIC, 2013; ISB, 2013; DHHS, 2012). For example, the Directorate of Health in Iceland has documented the processes to follow to encrypt and decrypt ID numbers; how to calculate age from an ID number; and how to manage re-identification keys. Such documentation assists in ensuring that staff conducting data de-identification will do so consistently and to the organisational standard. A few countries noted, however, that keeping data encryption algorithms confidential is essential to protecting the data and that, for this reason, encryption processes are not documented. The NBHW in Sweden, for example, maintains documentation internally but does not disseminate this documentation externally.

The use of pseudonyms to replace direct identifiers

Pseudonyms are created from direct identifiers for all key national health datasets in the United Kingdom (Scotland and Wales), Korea, New Zealand, Singapore and Czech Republic and for most key national personal health datasets in Japan, Finland, Netherlands, Norway, Denmark, Iceland, Sweden and the United Kingdom (England) (Table 7.1). This is a technique where identifying information about individuals, such as names, complete addresses and patient numbers are converted to a meaningless name or number in a consistent manner. The consistency of the application of the pseudonymisation algorithm permits record linkage among databases for approved projects. Such approaches to dataset linkages provide a higher level of protection of data subject's privacy. This is because statistical data linkers have no need to ever see the direct record identifiers to conduct an approved data linkage.

The assignment of a pseudonym may be done it a way that permits it to be reversible or not. When a pseudonym is reversible, then a process is needed to separate the key to its reversal from the dataset and to secure the storage of the key. Key storage is extremely helpful to be able to replicate and validate study findings at a later point in the future, to correct the original dataset for data quality problems identified by researchers, and to be able to extend a study to include a linkage to new years of data or to include a new dataset. Key storage also contributes to the efficiency of data linkages, as it permits keys to be assigned once and then reused to established links for future approved projects.

Creating pseudonyms

Country experts provided descriptions of the processes undertaken to create a pseudonym to replace identifying information within datasets (Canada, Denmark, Finland Iceland, Japan, Korea, Singapore, Switzerland, and United Kingdom). Often country experts indicated that commercial software provided the tools that they used to encrypt data to create pseudonymised ID numbers.

In Denmark, for example, the personal identification number (PIN) is encrypted using a hashing algorithm. The PIN carries information within in, such as the exact date of birth, which is then concealed through the encryption exercise. The same algorithm is used for all registries so that the pseudonymised ID number can be used to link datasets for approved projects.

A set of identifying variables may be included in the creation of the pseudonymised ID number. In Switzerland, for example, the Federal Statistical Office creates a pseudonym from the patients' first and last name, sex and data of birth. The same algorithm is used to create the pseudonym for all datasets and, as a result, the pseudonym is used to conduct deterministic record linkage.

The United Kingdom (Wales) provided an example of the use of a trusted third party to encrypt the identifying variables. Within the UK Wales SAIL project, datasets from government, other agencies and primary health care clinics are divided into two components by data custodians so that the identifiers are stored in one file including names, addresses, full postal codes, sex and complete birthdates and the content of the dataset in a second file. The datasets are transmitted from custodians to a trusted third party (NHS Wales Informatics Service). The third party compares the identifiers against the dataset of patients registered with a general practitioner to clean and correct the patient identifying number (NHS number). The third party then encrypts the NHS number and includes the encrypted ID within the datasets containing the clinical content.

When encryption algorithms are used to create a pseudonymised ID number, there can be a subsequent future need to re-identify the data. Reasons typically include replicating or extending approved studies, but may also include extensions involving contacting data subjects. The ability for future researchers to replicate a study's findings is an important factor in ensuring research results are valid. Experts in Singapore and Finland emphasised the importance of a reversible pseudonymised ID number so that data could be re-identified for an approved reason.

In Finland, THL explained that all data custodians involved in a data linkage study agree which one of them will hold the key to re-identify the data involved in a data linkage project. THL would typically ask them to retain the re-identification key for project data for up to five years, which would provide time to enable the data to be re-identified during the study period, if it became necessary to do so.

Evaluating and addressing data re-identification risk

Several countries reported that there are processes in place to evaluate the risk that key national datasets, after application of data de-identification steps, could become re-identified (Table 7.1). Countries reporting that all key national health datasets are evaluated for re-identification risks included the Czech Republic, Singapore, the United Kingdom (Scotland and Wales), Italy, the United States, and New Zealand and for the majority of datasets in Canada, Turkey, Finland, Spain, Sweden, the Netherlands and the United Kingdom (England). As was discussed in the previous chapter, some countries approve the transfer of de-identified microdata to third parties for approved projects, such as other government ministries, the academic or non-profit sector or the commercial sector. Whenever such transfers are possible, the need to evaluate re-identification risks is greater.

Two particular areas needing evaluation were raised by country experts. The first is related to the security of the data encryption algorithm and its potential to be used in unauthorised data linkages and the second is the risk posed by variables within datasets that can be used to indirectly identify data subjects.

Protecting encryption algorithms

Risks associated with the consistent use of the same encryption algorithm for multiple datasets and across time, include that data holders could conduct unauthorised data linkages, or could attempt to hack the encryption algorithm. Several countries provided examples where steps have been taken to reduce this risk.

Canada (CIHI), Switzerland (FSO) and Korea (HIRA) provided examples where the encrypted ID number is replaced with a meaningless anonymous study-specific number whenever data are released to a third party. In Switzerland, when a data request is approved involving sharing de-identified record-level data from the FSO with an external party, the pseudonymised identifier is first converted to an anonymous study number. The algorithm to generate the study number is different for each dataset and changes for each dataset every year. The FSO maintains a mapping of the study numbers to the pseudonymised identifiers and, in so doing, is able to reverse the study number if there is, for example, an approval to extend a study to add a future year of data. Similarly, in Korea, HIRA uses a study-specific encryption algorithm and in Canada, CIHI uses a meaningless study-specific ID number.

An alternative practice was shared from Iceland. In Iceland, at the Directorate of Health, each dataset has its own encryption algorithm. The key related to each dataset is then stored so that the dataset can be re-identified for an approved use, such as a new data linkage project. The process of encryption and decryption is documented, as is the process for storing re-identification keys.

Masking indirect identifiers

While some data de-identification methods only concentrate on protecting against disclosure of direct identifiers, other methods involve modifying a wider range of dataset variables in order to reduce the likelihood that these variables could be used to re-identify the data. Applying such techniques can be referred to as data masking. Practical guidance and case studies applying more sophisticated de-identification approaches are available (El Emam et al., 2014).

Suppression, which is the complete removal of information from a dataset, usually the cutting of a dataset variable, is a common data masking technique. An example of suppression is the removal of variables containing names and postal codes from a dataset. Suppression can also involve dataset records, such as supressing records for particular individuals because the record is considered too unique, such as supressing a birth of quintuplets if there was only one such birth.

A variable with high re-identification risk may be valuable to retain the dataset. In this case, other techniques may be used such as releasing only partial dates, such as month and year but not day, or larger geographies, such as postal code districts rather than exact postal codes. Continuous variables may be top- or bottom-coded. In this approach, a variable may be allowed to be continuous throughout a certain range but no values can go above or below a pre-set limit, such as an upper and lower limit on household income. Variables may be aggregated into groups, such as age groups or disease groups. Sometimes "noise" is added to the data, such as swapping values among dataset records with similar characteristics, rounding values, or otherwise distorting the original values in a random manner. Interestingly, the addition of noise has not been found to prevent sophisticated data re-identification attacks and is not recommendable for data privacy protection (El Emam, 2013). Noise addition also creates problems for the validity of analytical findings from the data.

Experts in several countries provided examples of additional steps taken to de-identify datasets after direct identifiers have been encrypted or supressed. These steps typically involve identifying variables that could be combined to provide enough information to potentially identify a data subject, such as exact dates and exact locations. The United Kingdom (England) Health and Social Care Information Centre (HSCIC) reported typical examples such as restricting dates of birth to month and year or to single year of age and restricting postal codes.

Iceland's overall population is small and therefore the risk of re-identification of patients with rarer health conditions is often high. In Iceland, health data are often treated to reduce re-identification risk by providing, for example, birth month, age or age groups instead of exact birthdates. Further, sometimes only part of the postal code is provided to avoid identifying areas that are too small. If the exact birthdate is essential to a study, location of treatment may be aggregated to broad categories, such as primary, secondary or tertiary care. The de-identification steps are generally applied when the target study population is small or when the dataset will have a lot of variables within it which could make it easier to re-identify individuals.

In the UK Wales SAIL project, a trusted third party de-identifies the data including masking potentially identifying variables, such as replacing date of birth with week of birth in the dataset or replacing postal code with a variable identifying small areas with at least 1 500 residents.

The US AHRQ also provided an example of the use of a trusted third party to de-identify data. A national law was introduced to enable the AHRQ to be a safe harbour for the collection of data about adverse events occurring in hospitals. The law encourages providers to report adverse events without risk of their identity being revealed, which would expose them to potentially facing a law suit. AHRQ works with an external supplier to de-identify the data before it is transferred to AHRQ. The third party ensures that the de-identification requirements of the main US health privacy law (HIPAA) are met and that data are aggregated to a level where there are no small cell counts.

National experts reporting data masking techniques were asked if their organisations were using automated tools to assist with masking indirect identifiers. All countries indicated that they either proceed manually or have developed in-house tools to assist them. None are using commercial software to support this task. The US NCHS is investigating the use of commercially available tools to review datasets and quantify the re-identification risk they hold; and is investigating options to increase the automation that can be applied to de-identification. In Ontario, Canada, the Institute for Clinical Evaluative Sciences is using a commercially available tool to review datasets for re-identification risk before datasets are made available to approved researchers within a secure data access environment.

Public-use microdata files

A few countries de-identify microdata to the extent that it is considered safe for dissemination to the public or to applicants meeting limited approval requirements.

In the United States, the NCHS creates public-use microdata files from population health survey data. The surveys are a sample of the population, so there is doubt regarding who might be part of the sample. The survey data are treated by first removing direct identifiers and then applying data masking techniques to variables remaining in the dataset that may pose a re-identification risk. A trial and error process involving a large number of cross-tabulations of the data are undertaken to determine the variables that require masking. Masking techniques used include top and bottom coding the values of variables (such as income), grouping variables (such as geography) or altering dataset values to add noise, such as swapping blocks of values among records. This effort is labour-intensive and time consuming and more of an art than a science. A similar process is also followed by Statistics Canada whenever a public-use microdata file is created from health survey data.

In a second example from the United States, the AHRQ prepares public use microdata files from survey data by removing direct identifiers and detailed geographic identifiers. The geography provided is a state identifier and an urban/rural flag. Birthdate is removed and age group is provided. De-identification for health care providers is more challenging because sometimes there is only one provider per state for certain specialised services, such as organ transplantation. Rather than supressing or masking data, this risk is mitigated through requiring applicants for access to the file to sign a data sharing agreements that binds them to protect identities and to not disclose small cells. Such agreements are among the data governance practices discussed in the next chapter.

In Canada, CIHI has developed a public use microdata file that is a 10% sample of the national in-patient hospitalisation dataset (DAD). This file is available to university researchers and students through the Statistics Canada Data Liberation Initiative. Canada's population health surveys are also made available as public use microdata files through the same initiative. In Korea, HIRA makes available to external applications a sample of claims data that have been de-identified.

In Japan, the Insurance Bureau offers access to a sample of 1% of one month's outpatient claim records and 10% of one month's inpatient claim records. Direct identifiers would be encrypted and geographic identifiers would be supressed. Sampling provides additional protection against the risk of re-identification and the sample could be shared with approved researchers for use in rooms with more relaxed security than the standard the Bureau usually requests. This would enable researchers in smaller universities that cannot provide a secured room to have access to data. This would also offer researchers an opportunity to do preliminary analysis to prepare statistical programmes before accessing the secure room to run their final analysis on the full file. The sample would not be an appropriate method for rare health conditions, as data de-identification risks would be too high.

Micro-aggregate data

In between microdata and aggregated data are data that are very finely aggregated. Such data can be used by researchers to construct custom tabulations and may even support certain statistical modelling applications.

At the Federal Statistical Office in Switzerland, processes are under development to anonymise data to a degree that it could be made publicly available and could help the public to find their own answers to what are now frequent information requests. A possibility is a fully anonymised data cube that could be queried. Once this work is completed then the FSO may consider the development of a restricted space for access to higher risk data.

In the United States, the NCHS is developing a new on-line data query system for the National Health Interview Survey that will enable the public to submit queries for data tabulations in real time. Unlike the Andre remote data access service, which is targeted to experienced researchers, the new system will be useable by the general public. A statistician has been hired to explore the extent to which the NCHS could develop micro-aggregate data.

Weighing data de-identification techniques against dataset utility for the intended purpose

Fourteen countries indicated that for most or all key national health datasets the usefulness of the data for the intended project was evaluated when finalising data de-identification methods (Table 7.1). This step is very important to be able to realise the intended public benefit from a proposed project.

Data suppression and masking techniques can have detrimental impacts on the ability to conduct certain studies or on the validity of study findings when the data are analysed. In the UK Wales SAIL project, while indirect identifiers may be aggregated, further data perturbation is avoided, such as altering variable values or swapping data among records.

This type of treatment increases the occurrence of bias in research study findings and in statistics and it risks interfering with the discovery and analysis of rare events. To mitigate the resulting higher dataset re-identification risk, access to de-identified data are provided within a secure environment.

In the past, under the *Statistics Act*, Statistics Finland could only provide researchers with access to data that had been both de-identified (removal of direct identifiers) and had been masked to reduce the risk of re-identification. When data from THL was to be linked with data from Statistics Finland and then shared with an approved applicant, the data masking process of Statistics Finland was found to have a negative impact on analytical findings. Since September 2013, a revision to the *Statistics Act* permits a linked file with both THL and Statistics Finland data to be shared for analysis with only de-identification (removal of direct identifiers) and without any further masking.

Several countries reported that decision making about the treatment of variables that may indirectly lead to re-identification of data are conducted on a case by case basis. In so doing, the data risks and utilities of the case can be considered.

For example, in Finland, THL indicated that variables may be supressed or aggregated but whether or not this is done depends on the research project and whether or not the exact dates, for example, would be needed for the study to be valid. There is more concern about re-identification risks and therefore the treatment of potentially identifying variables for studies of health conditions that are relatively rare within the population. An example of a dataset with more re-identification risk concerns would be a study of artificial reproductive technology where the dataset contained only one case of triplets born.

In a final example, in Korea at HIRA, there is not one methodology applied to reduce re-identification risk, but rather the treatment depends on the nature of the project proposed. The HIRA review committee evaluates the re-identification risk associated with a proposed project and recommends how potential indirect identifying variables will be treated.

Considering the broader data security environment when deciding on de-identification methods

The United States and the United Kingdom provided examples of how decision making about data de-identification processes are taken within a broader context of other measures that have been put into place to protect data privacy and security.

In England, the Health and Social Care Information Centre (HSCIC) allocates data into tiers according to its risk to patient's privacy. At the lowest tier are statistical publications and ad hoc tabulations which present aggregated data or results of models and pose the least privacy risk. For these data, a template is used to document the nature of the data release and what could possibly be revealed. A Small Numbers Panel of internal experts may be convened to review the risk of disclosure of data subject identities whenever there is a new publication or a more complex tabulation to be released. The next highest risk is de-identified microdata where a risk analysis has determined that the re-identification risk is low. In this case, the microdata could be made publicly accessible. The next higher level of risk is record-level data where direct identifiers have been removed and ID numbers pseudonymised but where combinations of variables create a re-identification risk. Such data are treated in a similar manner to fully identifiable data. Access to the data are controlled and individuals must apply to and receive approval from a designated body before being granted data access to the data. At the highest risk are identifiable data. Such data are only shared in compliance with the *Data Protection Act* and following the approval of a designated research ethics board.

In the United States, the National Centre for Health Statistics (NCHS) approach is to increase the effort to de-identify microdata wherever there is reduced security/supervision of data users. Microdata files destined for public dissemination receive the most treatment to de-identify the data to reduce the risk of violating the confidentiality of the data subjects and this treatment includes the use of variable suppression and masking techniques. Data masking techniques, which reduce the datasets utility and validity, are not applied to the data when the data can be accessed only within a secure, supervised environment. In this case, removal of direct identifiers protects data subject's privacy when the data are being analysed and the secure environment protects against the re-identification of the data.

Thus data de-identification rules must be developed in conjunction with an evaluation of the data governance overall including restrictions on access to and use of de-identified data. The next chapter presents data security environments in OECD countries and how they relate to the approaches to data de-identification that are taken.

Data de-identification practices that consider the "big picture": data protection, security and utility

Data de-identification methods are a moving target and practices that are satisfactory today will need to be revised with the introduction of new technologies, new health data and new data privacy protection risks. Thus the onus is on data custodians to keep abreast of best practices. A further challenge is that data de-identification is a specialised activity and it can be difficult for non-specialists to understand the strengths and limitations of various approaches. Documentation of data de-identification approaches is important to increasing understanding, as is including experts in data privacy protection in the development of data de-identification methods. Such collaboration creates potential for shared learning and the identification of best practices. Certain de-identification steps are critical and should always be followed to enable direct identifiers to be protected from disclosure and to enable indirect identifiers to be identified and treated to protect against dataset re-identification. At the same time, the steps followed need to accommodate future approved uses of data, such as dataset linkages. Guidelines on dataset de-identification should go beyond microdata to be inclusive of the potential risk of tabulated data or statistical models to indirectly reveal a data subject's identity. Lastly, to ensure that the data de-identification practices that have been planned for are undertaken, processes need to be audited or reviewed for compliance with standards.

The Advisory Panel of Experts on Health Information Infrastructure identified the following data de-identification practices as key to ensuring privacy-protective data use:

6. Best practices in data de-identification are applied to protect patient data privacy

Data de-identification practices

a) Document data de-identification methods.

b) Involve a data privacy expert in the development or review of de-identification methods.

c) Define direct and indirect identifiers.

d) Delete direct identifiers or, where necessary create a pseudonym from direct identifiers.

e) Replace the pseudonym with a meaningless study number when releasing any data to a third party.

f) Store the mapping between the identifiers, the pseudonym and any study numbers for future approved uses, such as informing data subjects and the replication, validation and extension of a study.

g) Include general rules for the treatment of indirect identifiers through data masking techniques.

h) Consider the impact of the treatment of indirect identifiers on the study results and consider other measures to mitigate privacy risks if the treatment of indirect identifiers will unduly damage the study findings.

i) Include guidelines on cell counts and indirect disclosure risks for tabulations and results of scientific research that are to be placed in the public domain.

j) Audit the data de-identification process to ensure all steps have been followed.

References

CIHI – Canadian Institute for Health Information (2014), "Privacy Policy on the Collection, Use, Disclosure and Retention of Personal Health Information and De-Identified Data", Canadian Institute for Health Information, Canada see https://secure.cihi.ca/free_products/Privacy-Policy_Revised-April-2014_EN.pdf, accessed 4 February 2014.

DHHS – Department of Health and Human Services (2012), "Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule", Department of Health and Human Services, United States, 26 November, see www.hhs.gov/ocr/privacy/hipaa/understanding/coveredentities/De-identification/guidance.html#standard, accessed 1 August 2014.

El Emam, K. (2013), *Guide to the De-Identification of Personal Health Information*, Boca Raton.

El Emam, K. and L. Arbuckle (2014), *Anonymizing Health Data: Case Studies and Methods to Get You Started*, Second Edition, O'Reilly.

Fraser, R. and D. Willison (2009), "Tools for De-Identification of Personal Health Information", Pan Canadian Health Information Privacy (HIP) Group.

HSCIC – Health and Social Care Information Centre (2013), "Guide to Confidentiality in Health and Social Care", Health and Social Care Information Centre, England, see www.hscic.gov.uk/media/12822/Guide-to-confidentiality-in-health-and-social-care/pdf/HSCIC-guide-to-confidentiality.pdf, accessed 4 February 2015.

ISB – Information Standards Board (2013), "Anonymisation Standard for Publishing Health and Social Care Data", ISB 1523, 25 February, see www.isb.nhs.uk/library/standard/128.

PrivacyAnalytics (2014), "The Twelve Characteristics of an Anonymisation Methodology", see www.privacy-analytics.com/de-id-university/white-papers/the-twelve-characteristics/, accessed 6 August 2015.

Chapter 8

Health data security and management practices

This chapter reviews the internal data security and management practices among health data custodians and how they assure data security is maintained when data is transferred to and accessed by external approved researchers and custodians.

Highlights

Sound data security practices are essential to meet legal requirements and public expectations for the protection of their health information. They ensure that the data held by national custodians are safe, that they are safe during any transfers and that they remain safe when they are shared with others.

Basic features of good governance within data processors include physical security, IT security, and secure channels for data transmission. Other basic features include a separation of duties, where only employees that need to see identifiable data to process it do so; signed obligations binding employees to protect data confidentiality; and regular staff training about their responsibilities for data security and confidentiality protection.

Several countries have made their data security processes transparent to the public by publishing policy statements or guidelines at either the national level or the level of national data custodians. Examples of published guidelines were provided by Canada, Denmark, Finland, Korea, New Zealand, Norway and the United Kingdom. A few countries also engage external experts to test their security with examples provided from Switzerland, United States and the United Kingdom.

Experts in 14 countries indicated that a signed obligation, such as a data sharing agreement, was used to legally bind data recipients to the rules to be followed to protect the data. Many such agreements place a time limit on how long data can be held by the third party before they are destroyed. Mechanisms to assure compliance with data sharing agreements include evaluations of data security environments before data access is approved and follow-up audits.

Countries universally observe that researchers have a strong incentive to comply with the terms of data sharing agreements because any misuse of data could damage their careers. Some countries have additional penalties. A fine or criminal conviction can be imposed for deliberate misuse of data in Korea, Norway and the United Kingdom, and among statistical authorities in Canada and the United States.

Secure research data centres and secure remote data access systems are viable alternatives to transferring person-level data to data requestors. The common feature of these mechanisms is that researchers are not provided a dataset to analyse within their own organisation. Instead, approved researchers must either physically enter a secure research data centre or digitally enter a secure remote data access system in order to analyse data. Secure research data centres are in use in Canada, Singapore, the Netherlands and the United States. Remote data access systems offering researchers with real-time service and the ability to conduct sophisticated data modelling with appropriate software are available in Canada (Ontario), the United Kingdom (Scotland and Wales), the Netherlands and the United States. Such an environment is undergoing pilot testing in Korea and is in development in Denmark.

Data governance practices begin with the requirements of and internal policies of dataset custodians. The participants in this study described the data security protections that are in place within their organisations to ensure their internal data security and to protect the confidentiality of the data they are in custody of. Most commonly noted is a separation of duties, where only a small staff with specific job requirements access identifiable microdata; followed by initiatives for staff training on data security and confidentiality protection; and physical security, such as secure networks, firewalls and threat assessments.

Data governance practices include data security practices that enable third parties to access data from custodians while minimising risks to data subject’s privacy. These practices include binding data sharing agreements or contracts; follow-up processes for conformance to agreements; supervised research data centres and secure remote data access systems; and civil and criminal penalties for data misuse.

In general there are two complementary approaches to protecting the privacy of data subjects when their data will be used for statistics and research projects. In the first approach, there is careful attention paid to the data itself and treatments are applied to the data to render it as anonymous as possible while still enabling high-quality research and statistics to be produced from it. This is technically challenging as was discussed in Chapter 7. In the other complementary approach, data de-identification is only one of the mechanisms and practices put into place to create a governance framework around the development of statistics and research with the data to ensure that the data are not misused and that the privacy of the data subjects remains protected throughout the process.

Countries were asked to identify the controls used to manage dataset re-identification risks (Table 8.1). Fifteen countries were able to identify one or more controls that are used to manage re-identification risks for all or the majority of the key national health care datasets. Examples of potential practices included in the questionnaire were supervised data access facilities, data security audits and penalties for misuse of data. Among the countries reporting controls are used, examples of controls provided by countries included limiting staff access to identifiable data; policies and guidelines including data de-identification standards; data sharing agreements and contracts to bind data recipients to follow data protection requirements; secure data access centres and remote systems; rules for minimum cell sizes in tabulations to avoid indirect disclosure of patient's confidential information; criminal penalties and/or fines for data misuse; and support and systems for patients to register complaints.

Table 8.1. Percentage of key national datasets where data security practices to protect data from re-identification were identified

	Data security practices to protect data from re-identification identified
Czech Republic	100%
Ireland	100%
Italy	100%
Korea	100%
New Zealand	100%
Singapore	100%
UK Scotland	100%
UK Wales	100%
United States	100%
Canada	88%
Japan	86%
Spain	75%
Netherlands	71%
Norway	70%
UK England	60%
Denmark	22%
Switzerland	20%
Finland	0%
Israel	0%
Iceland	0%
Turkey	0%
Sweden	ns

Source: Author's own calculations based on the results of this study.

There were also data security practices identified that place heavy restrictions on the use of and access to de-identified data for statistics or research in the public interest. These included practices that limit data access to aggregated data only (Italy), and policies that remove identifiers at the processing step rendering data linkage impossible (Japan).

Guidelines and policies to protect data privacy and security

The development and publication of policies or guidelines either at the national level or at the data custodian level greatly increases public transparency regarding the steps that are taken to protect health information privacy and security and provides a means to improve consistency within and among dataset custodians and a basis from which training courses and materials can be developed. International efforts, such as the standards and guidelines set by the International Standards Organisation regarding privacy and security requirements of EHR systems (ISO/TS 14441:2013); security of electronic health records communications (ISO/TS 13606-4:2009); and data protection to facilitate transborder flows of personal health data (ISO 22857:2011); support harmonisation of national data security and privacy protection practices. Country experts provided examples of the guidelines and policies that have been developed to protect data privacy and security.

In New Zealand, the legislative framework and the requirement for research ethics approval for identifiable data release are national in scope. Other policies are at the organisation level, with the Health Ministry having established its own internal policies. Consistency in policies among ministries is necessary to promote consistency among teams processing data. There was a project undertaken within the past year to develop guidelines and business rules for data handling. The need to develop formal guidelines arose as a result of a high profile agency that experienced a data breach. This raised the need to bring governmental agencies to a similar level of maturity of their internal systems. There is now a consistent data breach notification process across government. In the next 2-3 years there will be further development of information management guidelines. This work will support consistency among the different agencies in custody of health information. Further developments could include the appointment of a chief privacy officer for government who would take responsibility for data privacy, PIAs, data security and ICT systems. Other areas of work include the provision of access for citizens to their own data held by government.

The UK Information Commissioner's Office (ICO) has published a code of practice for data sharing (ICO, 2011). In England, for all of the NHS, there is a document that describes the NHS Anonymisation Standard. The HSCIC has written documentation regarding controlling re-identification risk when data are disclosed (HSCIC, 2013). The HSCIC statistical service also provides guidance to staff regarding disclosure control. The HSCIC conducts and documents privacy impact assessments for new data collections or projects involving personal health data.

In Denmark, the DPA provides guidelines regarding following the requirements of the national data privacy legislation. The SSI follows the DPA guidelines which provide guidance for all types of personal data. The unit providing research services has clear guidelines that they follow regarding data disclosure. For example, there are disclosure guidelines for minimum cell sizes for aggregated data tables to reduce the risk of indirect disclosure of a person's identity. There is a protocol to follow for reporting the discovery of a data breach that should be followed by all actors in the health sector.

In Korea, the Ministry of Security and Public Administration has produced guidelines for government agencies regarding the processing of personal data. HIRA also has internal guidelines specific to its own data holdings. There are guidelines regarding the review of applications, publication of approval decisions, processing timeliness, de-identification, data retention, security and data access fees. The guidelines do not apply to private sector holders of personal health data.

In Finland, each governmental institution maintains its own guidelines; however, in broad terms the guidelines would all conform to legislation. Within THL there is an effort to standardise guidelines across registries, however, there are differences among registries (such as the congenital malformations registry) that do require different rules. There are no guidelines in place regarding the reporting of a data security breach. The data protection anomaly should be reported to the chief security officer who will decide the severity of the case and may issue requests for action to both the owner of the data and the communications department.

In Canada, CIHI provides public access to its policies related to the protection of data subject's privacy and data confidentiality on its website (CIHI, 2014). Included in these, CIHI has a policy on the collection, use, disclosure and retention of personal health information and de-identified data, a policy on privacy impact assessments, a policy on staff privacy and security training, and security incident management protocol. Canada also has best practices guidelines that were developed as part of a health systems use project endorsed by the Conference of Deputy Ministers of Health (Health System Use Technical Advisory Committee, 2010). A joint report of Canada Health Infoway and the Information Commissioner's Office for the province of Ontario identifies essential data governance mechanisms, including de-identification and data security, to enable the secondary use of data from electronic clinical records (Cavoukian and Alvarez, 2012).

In Switzerland, as the interpretation of the law has recently been made clearer, a working group was established to develop guidelines for the FSO to ensure consistency in data protection practices throughout the organisation. The guidelines will cover procedures including data anonymisation, data linkage, management of re-identification risk, data disclosure etc. The working group is meeting regularly to develop the guidelines which will be submitted to the Data Protection Authority for approval. The intention is to make the guidelines available to the public. At this point the guidelines are general for the FSO but, it may also be decided to write specific guidelines for health data. The focus of the FSO has been on practices that reduce the risk of a data breach. There have not been any data breaches and the procedures to follow in the event have not been developed. Any illegal activity would be reported to the police.

In Norway, the National Patient Register (NPR) publishes guidelines, rules and regulations governing data security.

In Sweden, there is one guideline concerning disclosure of registry data. At the National Board of Health and Welfare (NBHW), the processing of requests is centralised in one unit and the employees of the unit meet together once per week to discuss the requests that have come it; to discuss complex requests; and to make consistent decisions. The legal expert takes part in the weekly meeting.

Spain reports guidelines within the Ministry of Health for the review of applications for access to data, the publication of approval decisions, data de-identification and data security. There is also an internal guideline on reporting a data security breach.

Data security within data custodians

Data custodians described a variety of practices that are in place within their organisations that support protecting the privacy of data subjects and the security of the data.

Granting staff access to data

A separation of duties is practiced in many organisations where employees that require access to identifiable data to fulfil their duties are the only members that are authorised to access data at this level of detail. Other staff members that require data for analysis receive de-identified datasets to work with.

In Iceland, the ministry described that the staff authorised access to identifiable data is a very small team and that all team members are very experienced.

In Finland, THL described that all staff seeking access to data are required to apply for data access privileges according to three levels (read, create and update/manage the structure of the registry) and must obtain an explicit permission for accessing person identifiers. In Canada, CIHI described that to be provided access to data, the staff request for access to data must be approved by their manager as well as the managers responsible for the datasets requested and IT services. An end date is specified or, if the data is needed on an on-going basis, then no end date is specified.

In Spain, there is restricted access to data to only staff requiring access and access is granted via an electronic signature. Software is used to track access to data.

In Korea, at HIRA, the patient identifying number is encrypted on all of the data files and the number is only decrypted when identifiable data are required, such as for an approved linkage. Employees exporting data or decoding identifiers need an approval.

In Singapore, at the Ministry of Health, a specific unit and selected individuals (such as the data custodians and the IT department) are approved to process and access identifiable data in the course of their work. Other staff members who engage in analysis only access de-identified data.

In Switzerland, the FSO explained a challenged encountered in the employee access process that was addressed. In the past, the process to provide access rights to datasets to employees was linked to the job position rather than to the individual. Sometimes, when individuals changed jobs they retained access rights to datasets that they no longer needed. The management of employee access to data was changed to ensure that tracking is by employee and that employees only have access to the data they need for their current position.

In Canada, CIHI has developed practices to ensure that staff access permissions are up-to-date. All approved data access to staff members is reviewed and approved once per year. Annual reports are sent to dataset managers to review staff access to data and to make corrections, such as removing access to data for a staff member who has changed jobs.

Staff access to data is more open to data custodian employees in Denmark. In Denmark, SSI employees are granted access to the datasets of the SSI and can link registries as required for their work. Most staff members have access to most registries.

Training staff about their data privacy and security responsibilities

Data custodians provided examples of the approaches taken to ensure that current and new staff members remain aware of their data privacy and security protection responsibilities.

In Iceland, the small Health Ministry staff involved in processing personal health data including dataset linkages is very experienced. There is no formal training process, but in the event of the introduction of a new staff member, experienced staff would train them.

When a new analyst joins the UK SAIL team in Wales, they are coached by a more experienced team member regarding the requirements for data governance and data protection, as well as on the SAIL data system.

In Switzerland, employees of the FSO are trained by their supervisors regarding their responsibilities for data security and privacy protection. There is a project underway, however, to introduce FSO internal training in conjunction with the development of organisation-wide guidelines on privacy and security.

In Finland, data linkage activities in response to external data requests are undertaken within one unit. This unit also fulfils internal data linkage requests, however, some data linkages for THL’s own research activities are undertaken within other units. All employees must pass an online test for basic security knowledge. Each new employee should be brought up to speed about relevant legislation and rules by their head of unit or a mentor and there are organisation-wide events to educate all staff about security concerns few times a year.

Data linkage activities are also not concentrated within SSI in Denmark. At SSI, new employees are given documents to read regarding their responsibilities to protect data privacy and confidentiality and it is mandatory for new employees to sign a paper attesting that they have read and will abide by the rules. A lawyer within SSI offers a course for new employees regarding their responsibilities under the law regarding data privacy protection. Similarly, Statistics Canada requires new employees to read their responsibilities and attest that the responsibilities have been read and understood and that they will comply with them. Statistics Canada also requires new employees to take on-line training on data confidentiality and security and their responsibilities and they must pass the course.

In Canada, CIHI has a training programme for all staff on their requirements to protect data security and the training guidelines are publicly available on the CIHI website. In Spain, the Health Ministry also ensures that staff members are trained on their role in protecting data privacy and confidentiality.

In Korea, HIRA employees are provided online training of privacy protection yearly, online ethics courses biyearly, and information security training biannually.

Securing and monitoring staff access to data

Country experts provided examples of practices put into place that help them to ensure that staff access to and use of data is appropriate.

In Denmark, SSI staff access to the data warehouse is logged including both the connection to the warehouse and the transactions that occur within the warehouse.

In Korea, at HIRA, there is a non-stop monitoring system that constantly tracks access to data and data use in real time. If the system detects an unusual access or use pattern, then a warning message is sent to the dataset manager and to staff in the security division.

THL in Finland explained that while access is logged, data misuse patterns are very difficult to recognise. Direct identifiers are usually hidden in datasets created for research, but a small number of persons have access to registries for statistical purposes and for creating research datasets. These persons are approved access to check information about a single person (i.e. verify valid and existing IDs or track history) or to grab all information for processing in statistical analysis software e.g. SAS or R. Usage anomalies are detectable and database accesses are traceable back to individuals.

Physical and IT security within data custodian offices

Country experts provided examples of the features of the physical security of their premises and the IT security of their information systems that help to protect the data they hold.

In Iceland, at the ministry, physical access to work stations is restricted by requiring the use of e-cards to open doors within the premises. Staff are only authorised to access areas of the premises required for their work. Visitors must be signed in and accompanied at all times. Computers lock when they are idle. These precautions are part of a suite of security steps to ensure that identifiable data never leaves the premises.

In Iceland, for a project enabling physicians to use a smart card to access a database of identifiable medications use information there is a risk assessment underway. The database is protected by multiple-layer network firewall. The assessment is testing the system's ability to withstand an external attack.

In the United States, at AHRQ, staff permitted to work with identifiable data must keep their offices locked. The small number of staff working with identifiable data must access these data on a secure network that has no connection to the Internet. If the staff member only has an occasional requirement for access to identifiable data, then they are required to access the data from within the secure data centre at AHRC.

Some organisations, however, wish to enable teleworking without adding risk to data security. In Canada, CIHI has developed a secure environment where staff may log in and work with de-identified microdata remotely in a secure manner, and where data cannot be removed. This provides a secure means for employees to work remotely.

At THL in Finland, all information systems under development are required to identify misuse scenarios and other security risks as part of national security guidelines (VAHTI) and there are risk modelling processes and tools available for significant development projects. At the THL, most registry data are stored in a relational database. All users have individual usernames and passwords and access is controlled using access roles. Person identifiers are encrypted and hidden from those who do not have appropriate access rights. All use of registry data are logged. The database system is located in a separate network segment protected by additional firewalls. A firewall around the IT system was also noted by HIRA in Korea.

External data security audits

In Switzerland, the FSO is using external experts to check their security and ensure that current security is safe from risk of external attack. The security that kept the FSO safe from attack ten years ago is unlikely to still keep the FSO safe today. Beyond the data, there is a need to ensure that algorithms used for pseudonymisation, for example, are secure.

In the United States, the data security of the AHRQ is subject to regular threat analysis conducted by an external contractor.

In the United Kingdom, the HSCIC uses a government accredited security firm to attempt to break the security of the HSCIC every six months. These penetration tests enable the HSCIC to ensure that the data remains safe from outside threats.

External data processors and cloud computing services

Three countries responding to this study noted engaging with external service providers for assistance with the processing of personal health data (United Kingdom, Spain and New Zealand). As health dataset volumes grow, with the development of data from electronic record systems and the storage of genetic and genomic data, the need for such services is expected to rise.

In the United Kingdom, the English HSCIC does contract out for data processing services to private sector suppliers. In this case, the HSCIC retains all of the legal responsibilities of data controllership for all of the data that is handled by these suppliers. Spain indicated that engaging with external service providers is not usual however, under contract and with very strict conditions to protect the data, such services have been used.

As is discussed in a later section, New Zealand and Australia have collaborated so that they can both use external cloud computing services available from providers in both countries to process their personal health data.

In Denmark, external companies are used to host services but not to process data. An example of a hosted service is remote access for employees to servers within SSI to enable telework. In Finland, all data processing, where data contains enough information to identify an individual, is done within THL. Some data processing using aggregated data or anonymised microdata are done using servers managed by partner organisations.

Protecting data during the transfer process

A number of countries reported security protections to prevent against data loss during the process of transferring identifiable data between organisations for approved projects. The use of secure internet portals for the transfer of datasets were noted by Canada, Denmark, Finland, Ireland, Korea, New Zealand, Singapore, Spain, Switzerland and the United Kingdom. Encrypting data and then sending data on a CD or USB were reported by Finland, Ireland, Korea, and the United Kingdom. The use of a courier or recommended mail to ensure that the correct recipient signs for the receipt of the data was noted by Finland and New Zealand. Some countries made note of using a CD or USB for data transfers, but did not indicate that the data were first encrypted. Some countries do not transfer identifiable microdata outside of their organisation. As is discussed in a later section, it is possible to provide access to data without transferring the data.

Country experts provided additional detail about the steps taken by their organisations to protect data when it is being transferred.

In Switzerland, the de-identified microdata transferred by the FSO are always encrypted and identifiable data are never transferred. Encryption is also used for any data flowing into the FSO. Sometimes the FSO receives unencrypted e-mail with sensitive data attached from hospitals. The data security in some hospitals seems weak. Data that are fully anonymised, such as aggregated data, are not encrypted for transfer by the FSO.

In Japan, in order to transfer data from the Insurance Bureau to a university with a secure data access room, the university provides an empty hard disk to the Bureau. The Bureau then provides a copy of the database on the hard drive to the university. The data are encrypted and the drive is sent to the university by registered mail. In the case where a researcher is approved access to a de-identified sample of the dataset, then a DVD is used. The data are encrypted and the DVD is sent to the researcher by registered mail.

In Denmark, the SSI uses a secure web portal to exchange data with regions. DVDs are used to share data with approved researchers and recommended mail is always used to ensure the DVD is delivered to the approved recipient. Data shared via DVD is often, but not always, encrypted.

In Spain, a secure web portal is used to transfer data among public authorities responsible for health care provision. When data are provided to external data requestors, usually only a sample of the data are shared and the data are encrypted before transfer.

In Finland, some data transfers take place via a secure web portal and, in other cases; CDs are sent using recommended mail. All data are required to be encrypted if they are sent through unsecure transport.

In Korea, only a sample of insurance claims that have been de-identified may be transferred from HIRA to external applicants. The files are transferred via CD without encryption.

Data sharing agreements or contracts

Experts in 14 countries indicated that a signed obligation, such as a data sharing agreement or contract, is used to legally bind data recipients to the rules to be followed to protect the privacy and confidentiality of the data for which they have been approved access (Canada, Czech Republic, Denmark, Finland, Iceland, Israel, Korea, Norway, Singapore, Spain, Sweden, Switzerland, United States and United Kingdom).

Elements of the rules that are included in signed documents include:

- use the data only for the purpose for which it was approved (Canada, Czech Republic, Denmark, Iceland, Korea, Norway, Sweden, United States)
- protect data confidentiality of data subjects and follow data disclosure rules (Iceland, Israel, Norway, United Kingdom, Denmark, Sweden, United States)
- respect national standards for data confidentiality protection (Norway, United Kingdom)
- provide custodian with copies of data findings or publications (Iceland, Israel, Japan, Korea, Norway, United States)
- time limit before data must be destroyed (Norway, Sweden, Switzerland, United Kingdom)
- do not disclose data to third parties (Czech Republic, Israel, Korea, Norway, Switzerland)
- penalties and disciplinary procedures for violations of the agreement (United Kingdom, United States)
- legal responsibilities when receiving identifiable data (United Kingdom)
- do not share remote data access permissions with a third party (United Kingdom)
- do not attempt to remove data from a research data centre (United States)

- do not make the data public (Switzerland)
- do not attempt to re-identify persons (Czech Republic, Singapore, United Kingdom)
- do not attempt to link or merge data with other data (United States)
- submit a data destruction certificate when data are destroyed (Canada)
- allow a follow-up data security audit (Canada).

In Japan, the signed contract between the researcher (data requestor) and the Minister of Health, Labour and Welfare also stipulates specific requirements of the requestor's university including the creation of a secure room with particular data security features and a mechanism to control access to the room and to monitor those entering and exiting it.

In Finland, the requirements of researchers are similar to those described by countries requiring data sharing agreements or contracts. Those approved access to data receive information about the laws, norms and requirements of the receipt of personal health data from THL. They must complete a declaration of confidentiality that binds them to not release the identity of individuals. After their research is conducted they are asked to destroy the data file, or to fully anonymise the data files such that the file would not enable direct or indirect identification of individuals. They also must also submit a notification of the end of research to THL and provide a copy of their published research results to THL. Clinicians, who are frequently the researchers involved, are also bound to protect the identities of individuals through their professional requirements. Data requestors are also asked to destroy the data at a specific time in the future. Published reports are to be provided to the data custodian after the study permission has ended.

Not all requests for data require a data sharing agreement in Denmark. Regions and municipalities must sign a licence agreement with SSI that enables them to access data on a per-year basis. Researchers sign a project-based agreement that is specific to their approved project.

In Canada, data transfers both to and from CIHI are typically governed by the terms of data-sharing agreements that are in place with provincial/territorial ministries of health and which cover all data flows to CIHI from within the particular jurisdiction. The data-sharing agreements set out the purpose, use, disclosure, security, retention and disposal requirements of personal health data provided to CIHI, as well as any subsequent data sharing that may be permitted. The agreements also describe the legislative authority under which personal health information is disclosed to CIHI.

Time limits and extensions

There are no time limits or data destruction dates in data sharing arrangements with health ministries in Israel and the Czech Republic. In the Czech Republic, if the researcher wishes to conduct a new research with the same data, they must reapply for permission for the new use.

Agreements with the HSCIC in England typically state that the data must be destroyed by the data requestor after three years. In other countries, time limits before data destruction are set on a case-by-case basis.

Two experts described what researchers can do when they have not finished their project by the data destruction date. In Finland, if the requestor requires an extension to the time limit, they must request it from the THL. The permission to use data in Finland is for a maximum of five years.

In Iceland, the researcher must request a time extension from the Data Protection Authority before an extension can be granted by the Directorate of Health. Time extensions require a less formal process and can be approved by the DPA following an e-mail exchange with the researcher. Researchers can also request an extension of a study to add additional years of data. Researchers can also request that the data are pseudonymised and stored by the Directorate of Health, so the data can be re-identified for a future approved use. This is possible because, under law, the Directorate of Health can preserve datasets created for scientific research.

Mechanisms to assure compliance with data sharing agreements

In England, an assessment of a data requestor's data security takes place before data are provided to them. The evaluation of data requesting organisations security procedures is particularly strict in cases where identifiable data has been requested (NHS numbers, exact dates of birth). There is an Information Governance Toolkit that guides the process of evaluating the risk level of data receiving organisations. Organisations with higher risk may still be approved for access to data, but they would be more likely to be targeted for a follow-up audit by an external auditor. The need to establish a set of standards and a process for the up-front assessment of receiving organisations' data security came about as a result of a breach of hospital episodes data from an external organisation that was found to have had poor internal data security practices that did not conform with the terms of their data sharing agreement with HSCIC. The HSCIC has a follow-up process to contact data receivers at the three year point to confirm that the data has been destroyed. If they need the data for a longer period of time, they must request an extension to the time limit. It is the HSCIC that is responsible for conducting audits of any data receivers.

In Canada, jurisdictional data-sharing agreements between CIHI and provincial/territorial ministries of health generally contain audit provisions. CIHI audits a proportion of the external third parties that have received data from CIHI. The purpose of the audit is to ensure that the third party is meeting or has met its contractual obligations as set out in CIHI's non-disclosure/confidentiality agreement.

CIHI also follows-up with third-party data recipients to ensure that their data destruction obligations have been met. Recipients must provide a certificate of data destruction back to CIHI. Data recipients are also required to certify on an annual basis that they remain in compliance with the terms of their non-disclosure and confidentiality agreement with CIHI. This activity helps to ensure that recipients of CIHI data remain aware of their obligations to keep CIHI data secure during the long time period between disclosure of data and data destruction. In cases where is the data are needed beyond the time frame originally indicated, the data requestors may apply to CIHI for an extended retention period.

In Finland, a recent change within THL has been the engagement of a staff member whose responsibility it is to follow-up with researchers who have received access to data in the past but who have not submitted a notification of the end of their research in order to verify the status of the project and take action if necessary. Most researchers with approved access to data are in public institutions in Finland and have legislative requirements to protect the data they hold. Commercial entities, on the other hand, may not have the same culture of data protection. The THL follow-up activities were motivated largely because of the need to ensure compliance by commercial entities. The THL can retract its approval and close a project if it has concerns about the recipient. The THL does not have the legal authority to conduct site visits or an audit of a data recipient but they can file a notification with the Data Protection Ombudsman if they have a concern resulting from their follow-up. The Data Protection Ombudsman has the authority to conduct a site visit or an audit.

In Iceland, the Directorate of Health does not have the role to follow-up with researchers to see if they are complying with the requirements of their agreement; however, the Data Protection Authority will do so. Similarly, in Sweden, the Data Protection Authority can to follow-up with the researchers to ensure that they are complying with requirements for data security. This follow-up responsibility is not in the hands of the NBHW.

In Switzerland, the FSO has no follow-up mechanism with data recipients to ensure that the data are destroyed when the destruction date arrives. In exceptional cases, a letter is sent from the FSO at the end of a project to ask if the data has been destroyed. There is also no follow-up process to review and assure compliance with data sharing agreements in the health ministries in the Czech Republic and Israel.

Korea (HIRA) and the United States (AHRQ) signalled that the provisions of their agreements with data recipients that require them to submit tabulations or publications for disclosure review have been difficult to enforce.

Penalties for non-compliance with the law and data sharing agreements or contracts

Wherever researchers are granted access to personal health data, countries universally note that these researchers have a strong incentive to comply with their legal obligations and/or the terms of their data sharing agreements because any misuse of data could affect their current and future applications for data access. For professional researchers, the damage to their career is a substantial deterrent.

Legal requirements to protect data privacy follow personal health data once it has been shared in New Zealand and the United Kingdom. In the United Kingdom there are three levels of enforcement powers that the Information Commissioner's Office (ICO) can exercise in cases of the misuse of personal data. The first level is that the ICO is legally authorised to compel an organisation to change its practices to comply with the requirements of *the Data Protection Act* (DPA). The second level can be applied when an organisation has experienced a data breach/data loss as a result of poor data security and disregard for its responsibilities under the DPA. In this case the DPA can impose a civil penalty of a heavy fine. The third level is to charge an individual with a criminal offence in cases where personal data have been stolen or deliberately misused for profit. Under law, such a conviction could result in a prison sentence but such a penalty has never been applied to this offence. The practice has been to impose a fine and to exempt the offender from obtaining a criminal record. The ICO is interested in stronger sanctions in the United Kingdom as a deterrent. In cases where there is a criminal conviction, the ICO has the legal authority to pursue the offender in civil court to recover all of the profits they may have earned as a result of the data misuse. Public complaints are an important way that the ICO hears of potential misuse of their data.

In the United Kingdom, for the SAIL project in Wales, the data sharing agreement specifies the penalties and disciplinary procedures for violations of the agreement. These include being reported to their management; having their access revoked; having any of their products within SAIL system rendered inaccessible to them; a lifetime ban from SAIL; and prosecution. The measures applied depend on the nature of the violation.

In the United States, researchers violating the terms of their signed agreement with the NCHS for access to data are barred from current and potentially any future access to data. If found to have deliberately made false statements in any matter within the jurisdiction of any Department or Agency of the Federal Government, the law provides for a punishment by a fine or up to five years in prison or both.

In Korea, under law, a person who reveals personal information that comes to his/her knowledge in the course of business or provides any third person with such information without due authority, a person who knowingly receives personal information for profit or unjust purpose, and person who corrupts, destroys, or leaks any third person's personal information shall be punished by imprisonment for not more than five years or by a fine not exceeding KRW 50 million.

In Norway, a fine or criminal conviction can be imposed for deliberate misuse of data.

In Japan, there is no penalty in law for a data breach but there can be a penalty associated with breach of contract. In the case of a breach of contract, the ministry will publicise the incident including the researchers name and the name of the researcher's university. At present, data access is not offered to private-sector organisations. It is uncertain that the penalty of potential publication of the name of the researcher and the researcher’s university would have the same deterrent impact on a private organisation as it does on a university.

In Switzerland, only a very small number of researchers have approved projects with access to record-level data from the FSO. The researchers are all well-known to the FSO and understand the conditions of their data sharing agreements. For some more complex cases, a meeting has been held with the researchers to discuss data protection requirements. The researchers are sensitised to the consequences of a data breach and know that their access to data would be cut off.

In Denmark, if the SSI suspects a data breach then the organisation that received data would be restricted from access to data for a period of time and the incident would be reported to the Data Protection Authority for investigation.

In the Netherlands, citizens can contact the Privacy Commissioner’s Office (PCO) with complaints regarding a suspected data breach. The PCO can investigate and where breaches are confirmed can levy penalties.

Data breach experiences

Only two countries shared examples of confirmed data breaches of personal health data. There was one example of a data breach involving a national health dataset. In that example, a recipient of hospital data from a national custodian was found to have not followed the data security and protection requirements specified in their data sharing agreement. In this case the data recipient was penalised by having their access to the data removed until they could demonstrate through a security audit that their data security protections were adequate.

The other two examples of a breach of personal health data both related to data while it was in the custody of a hospital. In the first example, there was a breach of the privacy of hospital patient data by hospital staff. In this case, staff without a need to view, viewed patient records. In the second example, a hospital was found to have been negligent in the protection of data on computers that were being disposed of. In this case, the hospital had not encrypted patient data stored on these computers hard drives and had not taken steps to remove the data stored on the hard drives before the old computers were sold to members of the public. This hospital was levied a heavy fine for their negligence by the privacy regulator for their violation of the national privacy legislation.

An expert in Denmark indicated that the government is working on ways to increase awareness within the health care sector about the importance of reporting data breaches and tracking data breaches. There is a need to increase awareness among health care managers and administrators in particular.

Alternatives to transferring data to third parties

Secure research data centres and secure remote data access systems are viable alternatives to transferring identifiable and de-identified personal health data from data custodians to third party data requestors such as other government ministries, university and non-profit researchers, commercial researchers, or to foreign researchers. These secure facilities are very effective at both broadening access to data for approved projects while at the same time reducing the risk that data could become re-identified or otherwise misused.

In both of these access mechanisms, the commonality is that researchers are not provided a dataset to analyse within their own organisation. Instead, approved researchers must either physically enter a secure research data centre or digitally enter a secure remote data access system in order to analyse data. In both mechanisms, it is not possible for researchers to download, print or otherwise remove data from the secure environment. All that is permitted to leave the environment are tabulations and statistical results that have been verified to not violate the confidentiality of data subjects. There are more countries moving forward with secure remote data access than are investing in secure physical premises for researcher access to data. This is likely because secure remote data access services have less overhead costs than do secure physical data centres.

In Japan, researchers applying for access to insurance bureau data must work for a university with a secure research environment or prepare a secure research environment. This includes a locked room for data access that only the approved researchers may enter that contains computers that are not connected to the Internet and do not permit USB keys or external hard drives to be used. Those entering or exiting the room must log themselves in and out. The Bureau conducts security audits of the universities to ensure that the secure room and procedures conform to Bureau requirements. In order to meet the needs of researchers within organisations that cannot afford to provide a secure room, the Bureau is considering offering additional secure rooms, one in the Eastern and the other in the Western part of the country.

The Ministry of Health in Singapore offers researchers with approved projects access to a secure area within the ministry – the microdata access lab. The lab provides approved researchers with access to de-identified health data. It is a supervised lab and access to the lab is controlled. Researchers are only able to remove from the lab aggregated data and their results leaving the lab are vetted by a staff member to ensure confidentiality of the data. A record is kept of who enters the lab and at what time for an audit trail. Some researchers do complain about the need to travel to the lab to access data and the ministry is investigating alternative data access mechanisms that would still enable the data to remain secure.

In the United States, the NCHS has one research data centre (RDC) and, in recent years, has been able to collaborate with the US Census Bureau to make an additional three research data centres available to researchers working with health datasets. These RDCs are all close to large research centres (Boston, North Carolina and Washington DC). These supervised RDCs are for researchers conducting projects requiring detailed de-identified microdata, including the most sensitive data holdings of the NCHS, such as survey data linked to Medicare and Medicaid data or research with genetic data from the NHANES survey. The data in the research data centre is raw (cleaned) microdata where direct identifiers have been removed.

The AHRQ in the United States also offers a research data centre for analysts requiring access to files with more detailed information, such as geography, that cannot be included within public use microdata files. The data centre is a closed environment where researchers cannot send files or otherwise remove data. It is supervised by a staff member who vets for confidentiality protection the outputs of researchers that they wish to remove from the centre. There are generally a small number of analysts using the data centre each week and many of them are students.

In the United States, the NCHS offers a secure real time remote data access service to domestic applicants for access to de-identified microdata for approved statistical research projects. The service is called Andre. As is the case for the Research Data Centres, Andre provides researchers with access to raw (cleaned) microdata where direct identifiers have been removed. In this system there are checks for attacks on the data by monitoring for the submission of multiple similar tabulation requests. Researchers accessing data via Andre cannot withdraw from the system the record level data and their research outputs are vetted for confidentiality protection before they can remove them from the system.

The SAIL project in Wales provides secure remote data access to its de-identified data. In virtually all cases, SAIL projects take place within the secure remote data access environment. The advantage for researchers is that they can access the SAIL system from their own offices. Only in exceptional cases would it be possible for a researcher to receive de-identified microdata from SAIL. The UK bio bank, for example, obtained informed consent from individuals for a data linkage and data transfer. As a result, it was then able to be approved to receive linked data from SAIL.

The SAIL system sets a researcher's scope within the gateway, ensuring that the researcher can only access the data they are approved to access. The system checks for unusual or inappropriate data querying or several unsuccessful log-in attempts which could signify an attack. The technical team does penetration testing of the SAIL system to assess its ability to withstand external attacks and to assess the security of SAIL firewalls. When the researcher has completed their analysis, a human data guardian reviews all of the tables and statistical outputs the researcher wishes to remove from the secure environment. As a rule of thumb, no results or tables can contain counts of less than five persons. Low counts must be referenced as being "less than five". In a typical case, the review of the outputs by the human guardian takes no more than one hour to complete. However, if a problem is identified, then there is a back and forth process with the researcher to revise their outputs.

Within Scotland in the United Kingdom, there are five health related safe havens including the national safe haven called the Scottish Health Informatics Programme at the NHS National Services Scotland. All of these safe havens provide data linkage services and access to de-identified health microdata to approved researchers through a secure real-time remote data access system with features similar to those just described for SAIL.

In Korea, HIRA is establishing a new remote access management system so that approved researchers can have secure remote access to the data. The new system is web based. The system will provide researchers with statistical software tools (SAS and R) that can be used to write and submit programmes in real time within the secure system. At present the new service is being pilot tested with 30 researchers. In most cases the system is fast but it does depend on the size of the data file being analysed and, for large files, the system can take longer to process the remote submission. Prior to introducing this new system, approved applicants were required to analyse the data within HIRA.

In Ontario Canada, the Institute for Clinical Evaluative Sciences (ICES) has established a new remote data access service in 2014 in order to enable linked and de-identified data to be used for approved projects by academic and non-profit researchers in Canada. Researchers are provided a user ID and password that enables them to access the approved de-identified record-level data in a secure environment over the web. Statistical software is available in the environment including SAS and R. No data, tabulations or statistical outputs can be printed or downloaded from the system. Researchers must request to have their results reviewed by an ICES staff member for confidentiality protection before results may exit the system.

In Denmark, the SSI is developing a remote data access service that follows the example of the remote data access service developed by Statistics Denmark. Applicants with approved projects could be approved for a digital signature. This signature is then used to access a secure environment over the web. Within the secure environment the researcher can use certain statistical software (such as SAS or STATA) to analyse record-level data. Record-level data cannot be downloaded from the secure environment. The only output from the secure environment is aggregated data or results of statistical models. The remote data access service of the SSI must be submitted to the DPA for approval before it can be launched.

Statistics Netherlands has a secure remote data access service. The service is only available within approved secure locations. Locked rooms where researchers with approved projects can access the secure data system of Statistics Netherlands are available at RIVM, other government institutions and within some universities. Where required, researchers may also be approved access to a secure room within Statistics Netherlands. Once in the secure room, researchers access a secure remote data access facility via a computer that requires fingerprint authentication. The facility prevents data from being copied or printed and the outputs from the research are submitted to be checked for confidentiality protection before a researcher can remove them from the secure facility. Health care TTP offers a secure remote data access service to de-identified microdata that is available to researchers via their own desktop. The security is higher for the Statistics Netherlands facility because Statistics Netherlands provides access for approved projects involving tax and social security data as well as health data. Researchers with approved projects must visit Statistics Netherlands to have their fingerprints taken. A catalogue describing the availability and accessibility of on-site and remote access to data from Statistics Netherlands including data security requirements and responsibilities is available in English (Statistics Netherlands, 2013).

Statistics Finland has introduced a remote data access service that can be provided to researchers working with files that have been linked to the holdings of Statistics Finland. The remote data access service provided by Statistics Finland is not in real time. It involves the researcher submitting programmes that are run by a Statistics Finland employee, the results are reviewed for protection of confidentiality and privacy and then results are provided to the researcher.

The Ministry of Health in Spain provides a remote data access service and a secure room within its facility where approved researchers may access a sub-set of the ministry's de-identified registries for approved purposes. Researchers that are approved to analyse the death index or who have been approved access to a cohort may access these services.

In Switzerland, for some data linkage requests where the requestor needs to link a cohort of data that they have collected to a dataset of the FSO, then the FSO may require the researcher to conduct their analysis within a restricted space inside the FSO.

Data security practices are essential to meeting legal requirements and public expectations

Data security and management practices are key to meeting legal requirements and public expectations for the protection of their health information. They ensure that data held by national custodians is safe, that it is safe during any transfers and that it remains safe when it is shared with others. Countries provided excellent examples of how data are kept safe through strong internal policies and guidelines and practices and a set of data governance mechanisms that provide a strong protection against re-identification and breach risks. These include secure channels for data transfers; data sharing agreements and contracts binding data recipients to the rules they need to follow to protect data privacy and confidentiality; mechanisms to ensure compliance with data sharing agreements and contracts including follow-ups and audits; and penalties for non-compliance. Further, countries provided examples of how access to microdata for approved projects can be provided without transferring data to third parties. These mechanisms are secure supervised research data centres and secure remote data access facilities. These secure mechanisms for data access are particularly promising for the future of national and multi-country statistics and research projects.

The Advisory Panel of Experts on Health Information Infrastructure identified the following data security and management practices as key elements of privacy-protective data use:

7. Best practices in data security and management should are applied to reduce re-identification and breach risks

Data security and management practices should provide for:

a) Controlling and monitoring physical and IT data security within data custodians and processors.

b) Controlling and monitoring to ensure that access to and use of personal health data within data custodians or processors is performed by staff subject to confidentiality rules/regulations.

c) Limiting data transfers to and from data custodians or processors to secure channels.

d) Requiring legally binding contracts with recipients of personal health data or de-identified person-level data from custodians or processors that specify the data confidentially and security requirements to be respected.

e) Ensuring data custodian staff, data processor staff and third-party data recipients of personal health data or de-identified person-level data have mandatory and periodic training on data privacy and security protection through on-line training or other means.

f) Before transferring data, reviewing the physical security and security policies and practices of data recipients and any parties mediating data transfers.

g) Conducting independent and random data security audits of data recipients and any parties mediating data transfers.

h) Following-up with data recipients to verify data destruction requirements and any other end of contract requirements have been met.

i) Offering alternatives to transferring data, such as providing data access within a research data centre or through a secure data portal, or analysing the data within a certified/accredited organisation.

j) Implementing penalties for data misuse by any party, such as contractual, financial or criminal penalties.

References

Cavoukian, A. and R.C. Alvarez (2012), "Embedding Privacy into the Design of EHRs to Enable Multiple Functionalities – Win/Win", www.infoway-inforoute.ca/en/component/edocman/resources/reports/338-embedding-privacy-into-the-design-of-ehrs accessed 15 April 2015.

CIHI – Canadian Institute for Health Information (2014), "Privacy Policies, Canadian Institute for Health Information", Canada see: https://secure.cihi.ca/estore/productSeries.htm?locale=en&pc=PCC30, accessed 4 August 2014.

Health System Use Technical Advisory Committee (2010), *Best Practice Guidelines for Managing the Disclosure of De-identified Health Information*, Data De-identification Working Group, Ottawa.

HSCIC – Health and Social Care Information Centre (2013), *A Guide to Confidentiality in Health and Social Care,* version 1.1, Health and Social Care Information Centre, England.

Information Commissioner's Office (2011), *Data Sharing Code of Practice*, United Kingdom.

NCHS – National Centre for Health Statistics (2014), "Data access agreement template", National Centre for Health Statistics see www.cdc.gov/rdc/Data/B4/AccessAgreement.pdf, accessed 4 August 2014.

Statistics Netherlands (2013), *Catalogue of Services: Microdata Services (On-Site/Remote Access) 2013/2014*, Centre for Policy-Related Statistics, version 1.

Chapter 9

The way forward for privacy-protective health information systems

This chapter presents national reflections upon past and future progress in the use of health data for health and health care quality monitoring and the barriers and facilitators to progress; recaps the data governance mechanisms proposed by the OECD in this report to support countries in maximising benefits and minimising risks from data use; and proposes future international collaboration to strengthen and harmonise privacy-protective health information infrastructure for national and multi-country statistics and research.

Highlights

Health data collected by national governments that can be linked and shared is a valuable resource that can be used safely to improve the health outcomes of patients and the quality and performance of the health care systems that serve them. Health Ministry leadership is necessary to ensure that delivering the data to manage this important sector is at the forefront of government policy and action.

In this study, countries were asked for their views about progress over the past five years in the use of personal health data to monitor health and health care quality and the outlook for the next five years. Eleven countries indicated that it has become easier or much easier to use personal health data to monitor health and health care quality over the past five years. Reasons for this included both technical improvements to data and data processing; as well as a strengthening of legislative frameworks governing health information privacy and greater clarity about the interpretation of legislation in practice. Sixteen countries are optimistic that they will be able to link datasets to monitor health and health care quality over the next five years and 13 countries indicate that it is likely or very likely that data will be extracted from electronic clinical records for this purpose. This optimism is either because such monitoring is already in place or because of improvements in data infrastructure including data quality, tools for data processing and progress in developing and standardising electronic health record systems.

A few countries described that the development and use of health data is becoming a government priority and that there are legislative reforms that are in planning or in process that are designed to help to address current limitations to the use of data for health and health care quality monitoring. Still, there are unresolved challenges that may limit progress in some countries, including uncertainty regarding the impact of the proposed Data Protection Regulation in Europe, a lack of government priority on solving data use challenges, and the need for more time to implement electronic health record systems before data from such systems could be analysed.

As was presented in this report, countries that have developed strong health data governance frameworks provided good examples of how data can be used safely to benefit society. From their experiences, the OECD brought forward data governance mechanisms to maximise societal benefits and to minimise societal risks from the use of health data. Each mechanism was introduced in a focused chapter of this report.

Best practices in data governance require continual assessment and renewal. This is because the volume, velocity and variety of health data is growing rapidly and the technologies used to communicate, process and store data are evolving. Further, legal frameworks continue to be renewed to reflect societal values and address requirements of a changing health information landscape. On-going collaboration among stakeholders in the development and use of health data is essential to developing balanced policy decisions that can reach the goal of maximising societal benefits and minimising societal risks.

International collaboration in this dynamic area is essential for information about best practices and lessons learned in health data governance to circulate widely; and to support movement toward common best practices so that multi-country statistical and research projects are feasible.

With OECD populations ageing and a rising burden of chronic health conditions; health systems will be under mounting pressure to improve care quality and co-ordination to ensure health systems are efficient, productive and financially sustainable. Data will be needed to assess and compare the effectiveness of therapies and services provided to chronically ill patients; to support re-designing and evaluating new models of health care service delivery; and to contribute to the discovery and evaluation of new treatments.

Health data collected by national governments that can be linked and shared is a valuable resource that can be used to improve the health outcomes of patients and the quality and performance of the health care systems that serve them. It makes ethical sense to use these data to their fullest potential within a governance framework that protects the privacy of data subjects.

OECD studies in 2011 and 2012 revealed that while all countries are investing in health data infrastructure, there are significant cross-country differences in data availability and use, with some countries standing out with significant progress and innovative practices enabling privacy-protective data use; and others falling behind with insufficient data and restrictions that limit access to and use of data, even by government itself.

Health Ministry leadership is necessary to ensure that delivering the data to manage this important sector is at the forefront of government policy and action. Effective collaboration between health ministries, justice ministries and data privacy regulators is essential if governments are to evolve toward a situation where societal benefits from data use are maximised and risks to society from data use are minimised. At the same time, government needs clear and open channels to engage with stakeholders in the development and use of data, so that data governance frameworks and practices reflect societal values and priorities.

This chapter presents national reflections upon past and future progress in the use of health data for health and health care quality monitoring; recaps the data governance mechanisms identified in this report as key to privacy protective data use; and proposes future international collaboration to strengthen and harmonise privacy-protective health information infrastructure for national and multi-country statistics and research.

Progress during the past five years

In this study, countries were asked for their views about progress over the past five years in the use of personal health data to monitor health and health care quality and the outlook for the next five years (Table 9.1). Eleven countries indicated that it has become easier or much easier to use personal health data to monitor health and health care quality over the past five years. In these countries there have been improvements in data quality and data standards (Canada, New Zealand, Spain and Singapore); in the use of a consistent patient identifier (Netherlands); in data timeliness (Iceland); in the population coverage of electronic clinical records (Spain); in the population coverage of key datasets (Czech Republic, Spain and United Kingdom); in centralisation of data processing (Denmark); and in data linkage processes (Denmark, Netherlands, United Kingdom and United States).

There has also been a strengthening of data governance mechanisms including legislative reforms to protect personal health data (Israel); clarity about data governance including the definition of de-identified data and the rules for data sharing (New Zealand, United Kingdom); and the introduction of a trusted third party to conduct data linkages and de-identify data (Netherlands). Ireland's Health Research Board is developing a proposal for a data governance model to enable data access, sharing, storage and linkage for health and related research.

Overall, countries were more optimistic about recent progress in 2013 than was the case in 2011. At that time, 30% of countries indicated that it became harder during the previous five years to use personal health data to monitor health and health care quality. This view was expressed in four countries in 2011 that also took part in the study in 2013. By 2013, none of these countries indicated that the situation had become harder.

Table 9.1. Views about progress in and the future of health data use

	Thinking about the PAST five years, would you say that it has become easier or harder to use personal health data to monitor health and health-care quality in your country?	Thinking about the NEXT five years, how likely is it that your country will be able to use linked data to regularly monitor any aspect of health care quality?	Thinking about the NEXT five years, how likely is it that your country will be able to use data extracted from electronic clinical records to regularly monitor any aspect of health care quality?
Canada	Easier	Likely	Very likely
Czech Republic	Neither easier nor harder	Likely	Very unlikely
Denmark	Easier	Very likely	Unsure
Finland	Neither easier nor harder	Unsure	Very likely
Iceland	Easier	Likely	Likely
Ireland	Neither easier nor harder	Likely	Likely
Israel	Easier	Likely	Likely
Italy	Neither easier nor harder	Very likely	Likely
Japan	Neither easier nor harder	Likely	Likely
Korea	Neither easier nor harder	Unsure	Unsure
Netherlands	Easier	Likely	Likely
New Zealand	Easier	Likely	Unsure
Norway	Easier	Very likely	Very Likely
Singapore	Easier	Likely	Likely
Spain	Much easier	Unsure	Very likely
Sweden	Neither easier nor harder	Likely	Unsure
Switzerland	Neither easier nor harder	Likely	Unsure
Turkey	Much harder	Unlikely	Very unlikely
United States	Easier	Likely	Very likely
UK England	No opinion	No opinion	No opinion
UK Scotland	Easier	Very likely	Very likely
UK Wales	No opinion	No opinion	No opinion

Source: Author's own calculations based on the results of this study.

A group of countries, however, indicate that there has not been any change over the past five years in the use of personal health data for health and health care quality monitoring. Reasons for this include that personal health data protection requirements have made using health data more difficult (Czech Republic); the institutional setting continues to make the use of personal health data difficult (Switzerland); and there have been no changes made to the legislative framework for data protection (Finland, Ireland, Italy). In Korea and Turkey, efforts to improve data quality and use continued throughout the period.

Outlook for the next five years

Most countries are optimistic regarding the likelihood of being able to monitor aspects of health care quality over the next five years through linking datasets and extracting data from electronic clinical records.

Data linkages are likely or very likely because data standardisation and linkage methodologies are improving (Canada, United States); electronic health record systems enabling data across care settings are developing (Canada) or are developed (Singapore);

the current key national health datasets support data linkages (Czech Republic); databases have been made linkage ready (Iceland, United States); tools to enable data linkages are growing (United States); data is more accessible for use (Denmark); an evidence-based and quantitative approach to health care quality and governance is a government priority (Italy, Korea); progress in the availability of national health insurance data has been made (Japan); there is an increasing range of information from regional data repositories (New Zealand); and indicators based on data linkages are already developed (United Kingdom).

The extraction of data from electronic clinical health record systems is likely or very likely because there are national projects in the planning stage to allow data extraction for monitoring (Ireland); there are more hospitals and providers using electronic clinical records (United States); there are improvements in data standardisation (Canada, Spain) and in interoperability including data models, binding methods and reference terminologies (Spain); it is a government priority to promote efficiency and quality in health care through the development of medical databases (Japan); data are already being extracted from primary care service records for quality monitoring and all health care records will be structured, including hospitals, by 2015 (Finland);[1] inpatient data from electronic clinical records are already being transferred from every hospital in the country in real time to the inpatient hospital database located at the Directorate of Health (Iceland); there is already extraction of data to monitor some aspects of quality (United Kingdom).

The strengthening of data governance mechanisms in some countries also supports the likelihood that data linkages and extraction of data from electronic clinical records could take place. The efforts include that a data governance and sharing bill in Ireland is expected to set data sharing and linkage principles for all public bodies including requirements around structure, project governance and security; a review of the current legislative framework is underway in Norway to address barriers to data use; a legal decree states that the electronic health record will be operational at the national level by June 2015 in Italy ensuring the interoperability of regional EHR systems; and a recent legal clarification regarding data governance, increased interest in health data from policy makers and the creation of a national academy of quality in health care to conduct research to improve quality all support progress in Switzerland.

Among the 14 countries that responded to both the 2012 study on the development of electronic health record systems and to this study, optimism about the likelihood that data from electronic health records will be used for health care quality monitoring has grown. Four countries that had been unsure or considered this data use to be unlikely now indicate that this data use is likely or very likely over the next five years (Korea, Netherlands, Spain and United States).

Policy and technical obstacles to progress over the next five years

Three countries indicated they were unsure about future progress in the development of quality monitoring via dataset linkages. The development of the *European Regulation on Data Protection* which has not yet entered into force is causing uncertainty with respect to how it may impact upon existing health information systems (Finland) and on data linkages and data anonymisation specifically (Spain).

A group of countries are unsure if they will be able to extract data from electronic health record systems to monitor the quality of care within the next five years, either because more time will be needed or changes in policy will be needed. The development of e-health is still in a very early stage in the Czech Republic and therefore it is unlikely that data from such a system will be available within the next five years; data from the electronic health record system is becoming more accessible in Denmark but it is uncertain if such data will be available within the next five years; private hospitals develop and use their own clinical records in Korea and standards are needed before data could be extracted to monitor the quality of care at a national level; progress in Sweden depends upon whether all clinical records will be structured in a way that is suitable for statistics; and electronic health records are widely used in hospitals in Switzerland but not in private practices and the law that will standardise the interoperability of records is in development but not yet established.

A further challenge emerged from a previous OECD study that found variability across countries in the take up of internationally agreed standards for data elements within electronic health records. It also found that there are key data elements that lack agreed international standards (OECD, 2013). The lack of agreed standards creates an obstacle to multi-country studies and to internationally comparable health and health care indicators.

Governance mechanisms supporting privacy-protective monitoring and research involving personal health data

Decision making about potential statistical or research uses of personal health data should be taken after considering both societal risks from the data use and societal benefits from the data use. Optimal decision making about potential statistical and research uses of data can only be achieved if there is an overarching data governance framework in the country that has itself been optimised to minimise societal risks from data use and to maximise societal benefits from data use.

As was presented in this report, countries that have developed strong health data governance frameworks provided good examples of how data can be used safely to benefit society. From their experiences, the Advisory Panel of Experts on Health Information Infrastructure brought forward seven data governance mechanisms to maximise societal benefits and to minimise societal risks from the use of health data. Each mechanism was introduced in a focussed chapter of this report.

The Advisory Panel of Experts on Health Information Infrastructure identified the following key elements of data governance supporting privacy-protective data use:

1. The health information system supports the monitoring and improvement of health care quality and system performance, as well as research innovations for better health care and outcomes.
2. The processing and the secondary use of data for public health, research and statistical purposes are permitted, subject to safeguards specified in the legislative framework for data protection.
3. The public are consulted upon and informed about the collection and processing of personal health data.
4. A certification/accreditation process for the processing of health data for research and statistics is implemented.
5. The project approval process is fair and transparent and decision making is supported by an independent, multidisciplinary project review body.
6. Best practices in data de-identification are applied to protect patient data privacy.
7. Best practices in data security and management are applied to reduce re-identification and breach risks.

Each of these data governance mechanisms was presented with a short set of specific dimensions of the mechanism. Accompanying the fifth item is also a proposed *Risk-Benefit Evaluation Tool* that project approval bodies can apply as a support to decision making about proposed data development and uses (Chapter 6, Table 6.2).

Next steps

Best practices in data governance require continual assessment and renewal. This is because the volume, velocity and variety of health data are growing rapidly and the technologies used to communicate, process and store data are evolving, including, for example, cloud computing services. This creates a dynamic environment where data re-identification and data security risks are evolving.

Further, legal frameworks continue to be renewed to reflect societal values and to address the requirements of a changing health information landscape.

On-going collaboration among stakeholders in the development and use of health data, including legal experts, regulators, statisticians, IT professionals, policy makers, researchers, providers and patients, is essential to developing balanced policy decisions that can reach the goal of maximising societal benefits and minimising societal risks.

International collaboration in this dynamic area is essential for information about best practices and lessons learned in health data governance to circulate widely; and to support common best practices so that multi-country statistical and research projects are feasible.

This study reveals several areas where international collaboration is needed, in particular to:

- support countries in developing the norms necessary for governments to certify or accredit data processors
- develop guidance for the implementation of project approval bodies
- ensure that there are sufficient agreed international standards for data coding and interoperability
- support countries to evaluate which national legal frameworks for the protection of health information privacy provide adequate protections to facilitate multi-country statistical and research projects
- review current practices in patient consent and in waivers to consent to reach a common understanding about mechanisms that are privacy protective
- review developments in data security risks and threats and mechanisms to address them
- explore mechanisms to engage the public in discussion about data and its governance to ensure that there is good public awareness of health data, the benefits of its use, its protection, and the rights of data subjects.

The Advisory Panel of Experts on Health Information Infrastructure identified the following practices as key to ensuring that data governance mechanism will remain relevant over time:

8. Governance mechanisms are periodically reviewed at an international level to maximise societal benefits and minimise societal risks as new data sources and new technologies are introduced

Periodic review is needed to:

a) Share best practices in data governance, including:

- review and develop norms for the accreditation or certification of data processors
- review and develop guidance for the establishment of project approval bodies
- review privacy legislations in OECD countries, compare similarities and differences, and create a list of countries sharing similar and adequate data privacy protection
- review current practices in patient consent and reach agreement on privacy-protective mechanisms to request/waive consent for research and statistics involving large health datasets
- review developments in data security risks and in software and IT processes to assist with risk mitigation
- review approaches to public consultation and public information about data uses, risks and risk mitigations.

b) Monitor national implementation of best practices in data governance, such as the eight key data governance mechanisms included in this report.

Note

1. The clinical utilisation and data acquisition of the National Electronic Patient Data repository started in 2014. The Repository is being constructed in stages, and therefore during the next five years more information will accumulate to the national electronic repository from the public and private sector. As use of the repository becomes more widespread, it is possible that the use of national EHR data for research purposes will be enabled through legislative means.

References

OECD (2013), *Strengthening Health Information Infrastructure for Health Care Quality Governance: Good Practices, New Opportunities and Data Privacy Protection Challenges*, OECD Health Policy Studies, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264193505-en.

Annex A

Health Care Quality Indicators Expert Group Advisory Panel of Experts on Health Information Infrastructure

Advice and guidance on all aspects of the study and its advice are being provided by the Advisory Panel of Experts on Health Information Infrastructure (APHII). APHII is a multi-disciplinary panel of international experts with backgrounds in health policy, research, statistics, law, privacy regulation, and information technology and includes representatives from government, academia, industry and civil society. The APHII has convened quarterly since April 2013 to develop the study protocol, develop the survey instruments, review findings and develop conclusions.

Table A.1. Members of the Advisory Panel of Experts on Health Information Infrastructure

Mr. Joseph Alhadeff	Chair of the OECD Information, Computer and Communication (ICCP) Committee and Vice President for Global Public Policy and Chief Privacy Strategist, Oracle Corporation, United States
Mr. Bruce Arnold	Assistant Professor, School of Law and Justice, University of Canberra, Australia
Mr. Suso Baleato	Civil Society Information Society Advisory Council (CSISAC)
Dr. Fabrizio Carinci	Professor of Health Systems and Policy, School of Health Sciences, Faculty of Health and Medical Sciences, University of Surrey, United Kingdom
Dr. Fred Cate	Professor and C. Ben Dutton Professor of Law and Director, Center for Law, Ethics, and Applied Research in Health Information, Indiana University, United States
Mr. Stan Crosley	Counsel, Data Privacy and Health Information Governance team, Government & Regulatory Affairs Practice Group, DrinkerBiddle, United States
Ms. Agnieszka Daval-Cichon	Policy Officer, Healthcare Systems Unit, Health and Consumers Directorate General, European Commission
Dr. ConcettaTania Di Iorio	Legal Consultant LL.M M.P.H., Serectrix s.n.c., Italy
Mr. Brent Diverty	Vice President, Programs, Canadian Institute for Health Information (CIHI), Canada
Dr. Khaled El Emam	Professor and Canada Research Chair in Electronic Health Information, University of Ottawa, Canada
Mr. David Evans	Group Manager for Business and Industry, Office of the Information Commissioner, United Kingdom
Dr. Ronni Gamzu	Director,General Hospital, Tel Aviv Medical Centre, Israel
Mr. Robert Gellman	Privacy and Information Policy Consultant, United States
Dr Unto Häkkinen	Research Professor, CHESS (Centre for Health and Social Economics), National Institute for Health and Welfare (THL), Finland
Dr. Päivi Hämäläinen	Director -Department for Information and Senior Consultant National Institute for Health and Welfare (THL), Finland
Dr. Poul Erik Hansen	Vice-President, Sector for Health Data and Research, National Statens Serum Institute, Ministry of Health, Denmark
Ms. Karolina Hanslik	Officer at European Commission DG SANCO (Directorate for Health), Belgium
Dr. Jacques Huguenin	Head of Health Care Statistics, Office fédéral de la statistique (OFS), Switzerland
Dr. Sun Min Kim	Director, Department of International Cooperation, Health Insurance Review and Assessment Service (HIRA), Korea
Dr Toshiro Kumakawa	Director, Department of Health and Welfare Services, Ministry of Health, Labour and Welfare, National Institute of Public Health, Japan
Ms. Denise Lebeau-Marianna	Lawyer, Baker & McKenzie SCP, France
Dr. Janet Murray	Caldicott Guardian, Information Services Division, National Health Service Scotland, United Kingdom
Dr. John Parkinson	Director, Clinical Practice Research Datalink (CPRD), Department of Health, England, United Kingdom
Mr. Tapani Piha	Head of Unit, eHealth & Technology Assessment Unit, DG SANTÉ (Directorate for Health, European Commission, Belgium
Dr. Patrick Romano	Professor of General Medicine and Pediatrics, UC Davis Division of General Medicine, University of California, United States
Dr. Chaiki Sato	Assistant Professor, Graduate School of Public Policy, University of Tokyo, Japan
Dr. Roxane Silberman	Directrice de recherche, Secrétaire Générale, Comité interministériel pour les données en sciences sociales, CNRS, École Normale Supérieure, France
Mr. David Smith	Deputy Commissioner and Director of Data Protection, Office of the Information Commissioner, United Kingdom
Dr. Lies van Gennip	Director, National IT Institute for Health Care (NICTIZ), Netherlands

Annex B

Health Care Quality Indicators Information Infrastructure Questionnaire

The Health Care Quality Indicators Expert Group (HCQI) Information Infrastructure Questionnaire serves a very important role within the overall information gathering plan for the Health Information Infrastructure project. It serves to describe the availability and use of personal health datasets within countries in order to monitor progress since 2011 and to explore dimensions of national data governance, accessibility and protection. Given virtually all countries have two or more different national authorities in the custody of key databases, the completion of this questionnaire requires most health ministries to seek input from other organisations.

The questionnaire is organised in three parts, with each part a separate worksheet. It includes a worksheet that provides a glossary of terms used in the questionnaire and a worksheet providing a set of record linkage case studies for reference.

Part A seeks information about the development and use of personal health data. It asks about datasets at the national level and at the level of regions/states or networks of health care organisations. It seeks information regarding sources and uses of data including data linkages, access to data, views and outlook for the future and contact persons for key aspects of data governance including: laws or regulations governing the use of personal health data; data processing centres; centralised project approval bodies; policies or guidelines and practices for data de-identification; and policies or programmes for open government health data.

Part B seeks information about two recent national projects involving analysis of personal health data and Part C seeks information about a recent multi-country study. Priority is to be given to projects involving record linkage of health care data or extraction of data from electronic clinical records. Contact persons are requested for each project that we can then approach for a follow-up interview.

The questionnaire was developed with input from the APHII and distributed to countries in June 2013 for completion by September 2013. The deadline was then extended to January 2014 to accommodate countries expressing both interest in the study and a need for additional time to complete the survey.

We received completed questionnaires from 20 countries: Canada, Czech Republic, Denmark, Finland, Iceland, Ireland, Israel, Italy, Japan, Korea, Netherlands, New Zealand, Norway, Singapore, Spain, Sweden, Switzerland, Turkey, the United States and the United Kingdom (England, Scotland and Wales).

Table B.1. Countries that responded to the 2013-14 HCQI Information Infrastructure Questionnaire

Canada	Brent Diverty, Vice-President, Programs, Canadian Institute for Health Information
Czech Republic	Jan Alexa, Analyst, Ministry of Health of the Czech Republic
Denmark	Jonas Kähler, Academic, Statens Serum Institut
Finland	Päivi Hämäläinen, Director of Department for Information and Senior Consultant, National Institute for Health and Welfare (THL)
Iceland	Gudrun Audur Hardardottir, Project Manager, The Directorate of Health
Ireland	Gráinne Cosgrove, Statistician, Department of Health
Israel	Ziona Haklai, Head of Health Information Division, Ministry of Health
Italy	Ugenti Rossana, Director General, Health Information and Department of Planning and Organization of the National Health Service Statistical System, Italian Ministry of Health
Japan	Tomoyuki Kado, Section Chief, International Affairs Division, Ministry of Health, Labour and Welfare
Korea	Sun Min Kim, Commissioner for Healthcare Quality, Health Insurance Review and Assessment Service
Netherlands	Lies van Gennip, Director, National IT Institute for Health Care (NICTIZ)
New Zealand	Angela Pidd, Team Leader Data Management, Ministry of Health
Norway	Hanne Narbuvold, Director, The Norwegian Directorate of Health,
Singapore	Eng Kok Lim, Director, Performance & Technology Assessment Division, Ministry of Health
Spain	Arturo Romero Gutiérrez, Technical Advisor, Clinical Information Systems, Vice-directorate of Healthcare, Information and Innovation, Ministry of Health, Social Services and Equality (MSSSI)
Sweden	Max Köster, Project manager Register service, National Board of Health and Welfare
Switzerland	Jacques Huguenin, Head of Health Care Statistics, Swiss Federal Statistical Office
Turkey	İbrahim Doluküp, Head of Department, Ministry of Health
United States	Irma Arispe, Associate Director, CDC/National Center for Health Statistics
United Kingdom (England)	Candida Ballantyne, International Comparisons for Healthcare Quality Improvement, NHS Outcomes Analysis Team, NHS England, Department of Health
United Kingdom (Scotland)	Barbara Graham, Information Consultant, NHS NSS Information Services Division
United Kingdom (Wales)	Sarah Lowe, Senior Research Officer, Welsh Government

Annex C

HCQI Expert Interviews on Health Information Infrastructure

Interviews with contact persons identified through the country survey as experts in dimensions of data governance including legal frameworks for health information privacy protection, project approval processes, data security mechanisms, data access mechanisms. Experts were asked about the practices that are followed to initiate projects, approve projects, protect data security, provide access to data, supervise access to data, train staff and researchers, and assess results. They were also asked about efforts to centralise services for data processing, access and project approval and for their views on progress and the outlook for the future. Sets of interview questions tailored to the expertise of each type of expert to be interviewed and were developed with the advice of the APHII in September 2013. Interviews took place by telephone from November 2013 through to July 2014.

Table C.1. Participants to the 2013-14 HCQI Expert Interviews on Health Information Infrastructure

	Study participant	Title	Institution
Canada	Douglas Yeo	Director, Methodologies and Specialized Care	Canadian Institute for Health Information
Canada	Cal Marcoux	Chief Information Security Officer	Canadian Institute for Health Information
Canada	Anne-Mari Phillips	Chief Privacy Officer and General Counsel	Canadian Institute for Health Information
Canada	Mary Ledoux	Senior Consultant, Privacy	Canadian Institute for Health Information
Canada	Josée Bégin	Director, Health Statistics	Statistics Canada
Canada	Bob Kingsley	Assistant Director, Health Statistics	Statistics Canada
Canada (Ontario)	Michael Schull	President and CEO	Institute for Clinical Evaluative Sciences
Canada (Ontario)	Charles Victor	Senior Director, Data Platform,	Institute for Clinical Evaluative Sciences
Czech Republic	Jiří Holub	Director	Institute of Health Information and Statistics
Denmark	Jonas Kähler	Analyst	Statens Serum Institut
Denmark	Jan B. Hedemand	Acting head of unit	Statens Serum Institut
Denmark	Milan Fajber	IT	Statens Serum Institut
Denmark	Niels Berdin Flarup	Academic	Statens Serum Institut
Finland	Mika Gissler	Research Professor	THL National Institute for Health and Welfare
Finland	Arto Vuori	Development Manager	THL National Institute for Health and Welfare
Iceland	Gudrun K. Gudfinnsdottir	Project Manager	Directorate of Health
Iceland	Gudrun A. Hardardottir	Project Manager	Directorate of Health
Israel	Ziona Haklai	Head of Health Information Division	Ministry of Health
Israel	Talia Agmon	Head of Medical Ethics and Biotechnology, Legal Department	Ministry of Health
Italy	Claudia Biffoli	Director of Office IV - Directorate General for Health Information and Statistical System - Department of Planning and Organization of the National Health Service	Ministry of Health
Japan	Genta Kato	Bureau of Health Insurance, Office for Health Insurance System Enhancement	Ministry of Health, Labour & Welfare

Table C.1. Participants to the 2013-14 HCQI Expert Interviews on Health Information Infrastructure *(cont.)*

Study participant		Title	Institution
Korea (Republic of)	LIM Bong Hyun	Deputy Manager	Health Insurance Review & Assessment Service
Korea (Republic of)	SHIM Jae Yoon	Manager	Health Insurance Review & Assessment Service
Korea (Republic of)	KIM Kyoung Hoon	Senior Researcher	Health Insurance Review & Assessment Service
Korea (Republic of)	Hyun-Pyo Kim	Expert in Data linkages and Approval Processes	Health Insurance Review & Assessment Service
Netherlands	Michael van den Berg	Senior Researcher	National Institute for Public Health and the Environment, Netherlands
New Zealand	Simon Ross	Team Lead, Analytical Services	Ministry of Health
New Zealand	Phil Knipe	Chief Legal Officer	Ministry of Health
Singapore	Lim Eng Kok	Director, Performance & Technology Assessment Division	Ministry of Health
Singapore	Stanley Kok	Director, Legal Office	Ministry of Health
Singapore	Tong Ming Shen	Director, Health Information	Ministry of Health
Singapore	Kelvin Tan	Deputy Director, Regulatory Policy and Legislation	Ministry of Health
Spain	Maria Angeles Gogorcena Aoiz	Technical Counsellor, Healthcare Information Systems	Vice-directorate of Healthcare Information and Innovation
Spain	Maria Jesús Macias Fernández	Head of Unit, Development, IT Department	Ministry of Health
Spain	Manuel Cabrera Silva	Head of Unit, Infrastructure, IT Department	Ministry of Health
Sweden	Charlotta Sandström	Legal Expert	National Board of Health and Welfare
Switzerland	Jacques Huguenin	Head of Health Care Statistics	Swiss Federal Statistical Office
United States	Peter Meyer	Director	National Centre for Health Statistics
United States	Jennifer Parker	Special Projects Branch Chief	National Centre for Health Statistics
United States	Ernest Moy	Medical Officer	Agency for Healthcare Research and Quality
United States	Bob Gellman	Privacy and Information Policy Consultant	Private Practice
United Kingdom	David Evans	Senior Policy Officer, Public Services	Information Commissioner's Office
United Kingdom (England)	Chris Roebuck	Programme Manager	Health and Social Care Information Centre
United Kingdom (England)	Simeon Smith	Staff	Health and Social Care Information Centre
United Kingdom (England)	Andy Sutherland	Staff	Health and Social Care Information Centre
United Kingdom (England)	Paul Croft	Staff	Health and Social Care Information Centre
United Kingdom (England)	Dawn Foster	Staff	Health and Social Care Information Centre
United Kingdom (England)	Diana Paine	Manager	Department of Health
United Kingdom (Scotland)	Janet Murray	Caldicott Guardian, PHI Division	National Service Scotland
United Kingdom (Scotland)	Emily Jefferson	Co-Director	Farr Institute of Health Informatics Dundee
United Kingdom (Wales)	Kerina Jones	Associate Professor of Health Informatics, Swansea University	Secure Anonymised Information Linkage Project and Centre for Improvement of Population Health through E-records Research

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

The OECD is a unique forum where governments work together to address the economic, social and environmental challenges of globalisation. The OECD is also at the forefront of efforts to understand and to help governments respond to new developments and concerns, such as corporate governance, the information economy and the challenges of an ageing population. The Organisation provides a setting where governments can compare policy experiences, seek answers to common problems, identify good practice and work to co-ordinate domestic and international policies.

The OECD member countries are: Australia, Austria, Belgium, Canada, Chile, the Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Japan, Korea, Luxembourg, Mexico, the Netherlands, New Zealand, Norway, Poland, Portugal, the Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Turkey, the United Kingdom and the United States. The European Union takes part in the work of the OECD.

OECD Publishing disseminates widely the results of the Organisation's statistics gathering and research on economic, social and environmental issues, as well as the conventions, guidelines and standards agreed by its members.

OECD PUBLISHING, 2, rue André-Pascal, 75775 PARIS CEDEX 16
(81 2015 23 1 P) ISBN 978-92-64-24455-9 – 2015

OECD Health Policy Studies

Health Data Governance

PRIVACY, MONITORING AND RESEARCH

Contents

Consult this publication on line at ***http://dx.doi.org/10.1787/9789264244566-en.***

This work is published on the OECD iLibrary, which gathers all OECD books, periodicals and statistical databases. Visit ***www.oecd-ilibrary.org*** for more information.

OECD*publishing*
www.oecd.org/publishing

ISBN 978-92-64-24455-9
81 2015 23 1 P

9 789264 244559