For Official Use **DELSA/HEA/HCQ(2011)11**

Organisation de Coopération et de Développement Économiques
Organisation for Economic Co-operation and Development **09-Nov-2011**

English text only

DIRECTORATE FOR EMPLOYMENT, LABOUR AND SOCIAL AFFAIRS
HEALTH COMMITTEE

DELSA/HEA/HCQ(2011)11
For Official Use

Cancels & replaces the same document of 28 October 2011

Health Care Quality Indicators

SECONDARY ANALYSIS OF HEALTH DATA TO GENERATE HEALTH CARE QUALITY INFORMATION

DRAFT

Report prepared for the Health Care Quality Expert Group

Friday 18 November
OECD Headquarters, 2 rue André Pascal, 75016, Paris

OECD contact: Jillian Oderkirk, email: jillian.oderkirk@oecd.org, tel: +33 (0)1 45 24 76 03

JT03311190

English text only

TABLE OF CONTENTS

Tables

Figures

NOTE FROM THE SECRETARIAT

1. Based on input from the HCQI Expert Group, the October 2010 Ministerial Meeting and Forum on Health Care Quality included discussion of the importance of well-functioning national information infrastructure to improve indicators of health care quality. Such infrastructure needs to be systematic and efficient; capable of supporting linkages among data sources; and to provide appropriate protection of the privacy and confidentiality of health information.

2. In December 2010, the Health Committee endorsed further work to support development of health information systems to provide internationally comparable measures of health care quality.

3. In May 2011, the HCQI Expert Group agreed to undertake two reports supporting appropriate national information infrastructure development.

- A report on the potential, the barriers and the best practices in the linkage of personal health data for public health and health services research. The study involved a survey of country experiences and case studies of data linkage projects; and follow-up telephone interviews with country experts and linkage project leaders. The preliminary report is attached here. Please note that the case studies of data linkage projects (chapter 5 of this report) will be distributed at the meeting as a room document.

- A report on best practices in electronic health record (EHR) systems design and implementation that enable EHR data to produce health-care quality indicators. This project would also involve a survey of country experiences and follow-up telephone interviews with key individuals involved in electronic health record system design and implementation. This report would be prepared for the June 2012 HCQI meeting.

4. In 2009, the Health Committee endorsed further work on the benchmarking of health information technology. Such work required international funding support. The Commonwealth Fund and the Office of the National Coordinator for Health Information Technology of the United States have agreed to co-fund an international workshop to elaborate the indicators and processes to benchmark internationally the adoption and use of information and communication technologies (ICTs) in health. The workshop is in planning for January 30 and 31, 2012.

5. In June 2011, the Health Committee endorsed a proposal to hold a joint workshop of members of the Health Care Quality Expert Group and members of the Working Party on Information Security and Privacy to discuss enabling the secondary use of personal health data for public health and health services research. The aim of the workshop is to identify the privacy and confidentiality challenges, the barriers and the possible international actions. This workshop could be organised as a half-day meeting and take place on the morning of May 11, 2012, immediately following the meetings of the HCQI expert group and meetings of the Working Party on Information Security and Privacy. A new version of this report, reflecting the discussion of the HCQI Expert Group on November 18, could be a background document for this workshop.

6. Members of the HCQI Expert Group are invited to:

- Comment on the results of the study;
- Discuss preliminary recommendations and next steps for this study;
- Make recommendations for the continuation of this programme of work.

Secondary Analysis of Health Data to Generate Health Care Quality Information

Potential, Barriers and Best Practices in Data Linkage

EXECUTIVE SUMMARY

INTRODUCTION

7. Health data constitutes a significant resource in most OECD countries and it makes economic and ethical sense to use this data as much as possible to improve population health and the effectiveness and the efficiency of health care systems. Central to the assessment of both the health of populations and the quality and efficiency of health care services are data to measure, monitor and compare performance. Regional, national and international reports on health and health care are entirely dependent upon monitoring policies and investments in data infrastructure that either facilitate or restrict data and analysis (OECD, 2011).

8. Understanding the progress of the health of populations and understanding the performance and quality of health care systems requires the ability to monitor the same individuals over time, as they experience health care events, receive treatments, experience improvements or deteriorations in their health and live or die. It also requires understanding the distribution of health and health outcomes across different groups in the population and understanding variations in care quality and health outcomes.

9. This work has a few very important prerequisites. First, it is based on the study of real people in real-world settings, not specifically selected individuals in a clinical trial who do not represent the population. Thus, it depends on the existence of data representing the population within health care administrative databases, surveys and censuses. Second, it relies on a capacity to be able to follow the pathways of people in the population through different life events to measure change. This would include deaths following surgery, admissions to hospital following prescription medicine treatment, cancer survival, etc. Data to understand pathways and variations often requires the linkage of more than one database at the level of the persons within the databases because few databases have all the needed information.

10. The potential cost to countries of implementing new data collections for every new need for public health and health services monitoring is prohibitively high and wasteful of the existing information infrastructure that could respond to the information need with little additional cost. Further, many studies require a long-term follow-up, so introducing new data collections would not provide any answers for the benefit of the public's health today. Rising levels of chronic disease and multi-morbidity; concerns about the quality and safety of patient care; the need to measure and assure value for money for investments in health; and the need to allocate health system resources wisely are all too important to leave without good evidence for decision-making. This is particularly true when such evidence could be gathered from existing information through data linkages.

11. There are considerable differences across OECD countries in the extent to which personal health data may be collected, linked and analysed and the extent to which such data are currently contributing to monitoring population health and the quality of health care. The principle reason for avoiding data linkages

in some countries is uncertainty about legality of linkages, given existing laws that relate to the protection of health information privacy. The resources required to comply with legislative requirements to enable data linkages is a secondary problem, as is the cost of developing the technical capacity to undertake the work. A further problem is the failure of governments to communicate with the public about the availability of personal health data for research; the process in which it may be possible to access and use this data; and the benefits that accrue to the public's health of enabling research that is in their interest.

12. To be useful for the assessment of the quality of care, health and health care data collections need to be organised in a systematic and efficient way, and be structured to support linkage across data sources. At the same time, confidentiality of the data needs to be protected and privacy rights addressed (OECD, 2010). As improving information infrastructure requires both investment and legislation, improvements necessitate the support of national and federal governments. At the 2010 OECD ministerial meeting, health ministers concluded that further development of health care quality indicators was welcome and desirable and that such developments would require "better health information systems and more effective use of the data that are already collected" (OECD, 2011).

13. In May 2011, the OECD Health Care Quality Expert Group proposed undertaking this report which explores the challenges, the opportunities and the practices in the secondary use of personal health data for health research. The report focuses on the linkage of personal health records across multiple datasets within countries and across multiple countries.

Issues examined in this report

14. Chapter three reviews the information infrastructure within OECD countries that would support the analysis of personal health data for health and health-care quality monitoring and for research. This includes the extent to which data linkage studies are occurring to inform about health and health care outcomes and whether data linkages contribute to the regular monitoring of health care quality.

15. Section 3.1 reviews the extent to which each of the participating countries have national hospital in-patient data, primary-care data, cancer registry data, prescription medicines data, mortality data, formal long-term care data, patient experiences survey data, mental health in-patient data, population health survey data and population census or registry data. The challenges arising for some countries as a result of the need to negotiate data sharing arrangements across different data custodians are discussed in section 3.2. Section 3.3 reviews the infrastructure to support data linkages including the availability of unique patient identifying numbers and other identifying variables. In some countries, there is a strong infrastructure for health data linkages at the sub-national level, such as within states, regions or networks of health-care organisations. This capacity is discussed in section 3.4. Section 3.5 reports on the extent to which countries are conducting data linkage studies on a regular and occasional basis and the databases that are being used. It also discusses the extent to which there is regular monitoring of health-care quality via data linkages.

16. Chapter four discusses the protection of privacy in the collection and use of personal health data including data linkages. Section 4.1 notes the eight principles of data privacy protection from the OECD privacy framework and the legislations on protection of data privacy that have been enacted across the countries as noted by the study participants. An overview of how privacy principles are put into practice is presented in section 4.2 including how data linkage activities have been put into place that comply with legislative requirements; differing approaches to data de-identification; and the development of secure facilities for data with a high re-identification risk. Section 4.3 provides a review of the project approval process for data linkages in the countries and the specific case of researchers who request linkage of their own cohort of data to national databases. Data security is reviewed in section 4.4. This includes the security at the data custodians own facilities and the security related to researchers external to the data

custodian have received de-identified personal health data. Section 4.5 describes the particular challenges associated with multi-country projects and examples of success.

17. Chapter five provides summaries of recent projects involving the analysis of personal health data that were identified by countries because of their policy relevance. Section 5.1 presents case studies of data linkage projects that have contributed to the monitoring of the performance of the health system including the effectiveness, the efficiency and the safety of health care services and variations in health and health care outcomes within populations. A large multi-country study is also presented as are case studies of two very strong academic data linkage centres; and a new data linkage centre providing services to governmental and non-governmental researchers. Section 5.2 presents a summary of other projects identified by countries as important examples of the value to patient care and health policy of data linkages.

18. The views of the study participants about the strengths and limitations of their national data infrastructure and their ability to undertake data linkage studies and their outlook on the future are summarized in chapter six.

19. Chapter seven concludes with a set of preliminary recommendations for next steps at the international level to support the on-going development of national information infrastructure for data linkage and the protection of data privacy.

20. *Please note that chapter 5 will be distributed as a room document at the HCQI Expert Group meeting on November 18, 2011.*

Secondary use of personal health data defined

21. Health data is often originally collected for administrative purposes or for direct patient care. Re-use of this data for purposes other than those for which it was originally collected is considered a secondary use. Some of the most common secondary uses of health data include:

- Identifying the causes of disease, the prevalence of risk factors and identifying populations at risk;'
- Protecting public safety, especially with regard to infectious disease, but also in relation to prescription medicines, medical devices and environmental hazards;
- Needs assessment, monitoring and evaluation of services, with a view to providing an optimum performance of health care systems; and
- Improving the quality and safety of care in hospitals, practitioner's offices, clinics and other health-care settings.

22. Health data is personal when it is collected and stored at the level of individual patients or persons. Personal health data is needed to track events over time or across different health-care setting and to investigate the potential role of risk factors in the development of disease or the effectiveness of treatment. Often such analysis requires the linkage of personal health data across two or more data sets. Linkage occurs when records from the same patient, or the same person, in two or more different databases are merged together, creating a more complete health biography. 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 (See Annex 4: Glossary of terms).

23. Public registries, administrative databases and clinical records, including electronic health records, are all important sources of personal health data where analysis and dissemination of results are a secondary use of the data. Other important sources of health data include population and patient surveys and population censuses or registries.

Study method

24. A mail-back questionnaire sought information about the general environment in each country for secondary use of personal health data as well as specific case studies. Results were compiled into a database and statistical analysis was undertaken using Excel. The questionnaire was sent to the members of the OECD Health Care Quality Expert Group in July 2011 and responses were received from 15 countries in September and October 2011. Countries participating in the survey include Belgium, Canada, Cyprus[1,2], Denmark, Finland, Germany, Japan, Republic of Korea, Malta, Portugal, Singapore, Sweden, Switzerland, the United Kingdom and the United States.[3] Members represent the 35 member countries of the Organisation for Economic Cooperation and Development as well as a number of non-member countries who are participating actively in the HCQI project (see Annex A).

25. As part of this questionnaire, contact persons were identified who were knowledgeable about the general environment for secondary use of personal health data involving data linkages and multi-country studies. Experts with knowledge of national level studies, as well as regional, state and health-care network specific studies were identified. Structured telephone interviews were conducted with 28 selected experts from September 26 to October 18, 2011 (see Annex B).

NATIONAL INFORMATION INFRASTRUCTURE

26. National information infrastructure is quite strong across the countries participating in this study. All have the legal authority to collect identifiable personal health data and all are collecting identifiable personal health data at a national level. Countries also report no limitation in law affecting the retention of personal health information for their unlinked databases. All countries are legally able to analyse the data they have collected to monitor the public's health and to conduct research.

27. Many pursue data linkage studies on a regular basis and a number regularly monitor health care quality and the performance of their health system through data linkages. Challenges to pursing data

[1] Note by Turkey: The information in this document with reference to "Cyprus" relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognizes the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of United Nations, Turkey shall preserve its position concerning the "Cyprus" issue.

2 Note by all the European Union Member States of the OECD and the European Commission: The Republic of Cyprus is recognized by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

3 Italy participated in the telephone interview part of the study.

linkage studies, however, relate to multiple data custodians and the consequent necessity of the sharing of person-level data across different public authorities.

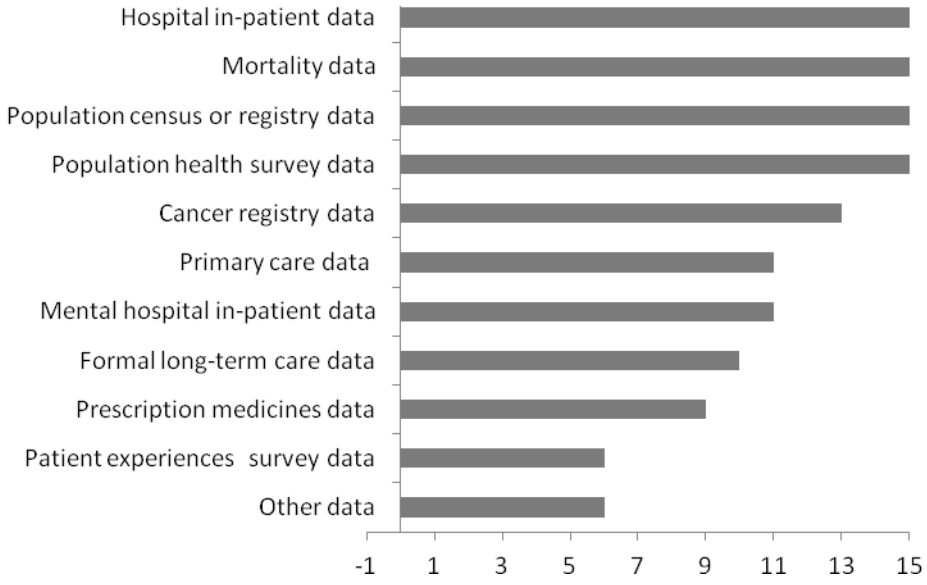
National databases

28. There is a strong underlying infrastructure for analysis of personal health data within the countries participating in this study. All fifteen participating countries have national inpatient hospitalization data, national mortality data; national population health surveys and a national census or a national population registry (See Annex C, Table C1). Thirteen have a national cancer registry; eleven have national data for primary health care and mental-hospital in-patient care; and ten have formal long-term care data. Less common are national data collections on prescription medicines (9) and patient experiences (6). Six countries have reported one or more other databases that are important to their national data infrastructure. These include emergency care data; clinical quality databases; data on births and congenital anomalies; retirement and disability pension claim data; disease management programme data; sickness fund data; dental care registries; and registries for diseases other than cancer.

29. All countries use their national databases to regularly report on health care quality (Table C2). Thirteen countries benefit from their inpatient hospitalization data to monitor health-care quality and twelve benefit from mortality data and cancer registry data for this purpose. Ten countries report using mental hospital in-patient data and population health survey data for health care quality monitoring. Eight countries monitor health care quality using primary health care data and seven use prescription medicines data; formal long-term care data; and population census or population registry data (in conjunction with health information). Five countries benefit from patient experiences data to monitor health-care quality and the same number also use other important databases to complement their programme of health-care quality monitoring.

30. Thirteen countries have national data at the level of individuals for mortality and in-patient hospitalizations (Table C3). Such data can be organised in a database where each row of the database represents an individual. This type of data is a prerequisite for detailed analysis of risk factors or determinants of health and health care outcomes and is a prerequisite for data linkage. Twelve countries have individual-level records in their cancer registry, population health survey data and population census or population registry data. Ten have individual-level data for primary care and nine have this data for formal long-term care and mental health hospital in-patients. Seven have individual records for prescription medicines and five for patient experiences.

Figure 1. Number of countries reporting national databases



Note: The data custodian should be a national authority and data should be included even when it does not cover 100% of the nation.

Source: OCED HCQI Questionnaire, Secondary Use of Health Data, September and October 2011

31. Countries were asked to report for all data available at a national level; even it does not cover 100% of the nation. While the impact of population coverage is minor in some countries, it can introduce significant biases in others. For example, some national databases in Canada are available for a limited number of provinces. In this case, the databases do not reflect the regional diversity of the country but do reflect the heterogeneity within the provincial populations. In the United States, national data on health-care encounters may be limited to particular sub-populations, such as individuals enrolled in Medicare (elderly persons) or Medicaid (lower-income persons) health insurance programmes or military veterans. In this case, the data is not representative of the underlying heterogeneity of the population.

Multiple data custodianship and data sharing

32. All countries report that there are several national government authorities, agencies or organizations acting as custodians of their national databases. National custodians include governmental departments or agencies responsible for healthcare or healthcare insurance; national statistical authorities; cancer registries; birth and death registries; national agencies responsible for health data collection or analysis; national authorities responsible for components of health care such as primary health care or care for veterans; university and scientific institutes; associations of local health authorities; and hospitals.

33. Denmark provides an example of data custodianship that is not at all atypical. Denmark reports that most health-care related national databases are in the custody of the National Board of Health, with the exception of prescription medicines that are in the custody of the Danish Medicines Agency. Population health surveys are in the custody of the National Institute of Public Health, patient experiences surveys are in the custody of the Capital Region for all of Denmark, while the population registry is in the custody of Statistics Denmark.

34. Some countries report further complexity in their national data infrastructure due to custodianship of national data at a sub-national level. For example, the United Kingdom reports custody of databases at the level of the individual countries within it and then, within each of the countries, multiple data custodians. The United States reports custodians of national data for particular sub-populations, such as military veterans or enrolees in Medicare and Medicaid insurance programmes.

35. The only exception to multiple custodianship is Switzerland, where the Federal Statistical Office is the single custodian of all of the national databases in their country's national health information infrastructure inquired about for this study.

36. What is important about multiple custodians is that, when they exist, there must then be legal frameworks and information custodian policy frameworks in place that provide for the possibility of the sharing of data. Without this, there is no possibility for any health or health care monitoring or research that requires person-level datasets from more than one custodian. Even when legal frameworks exist, data sharing can involve long and challenging negotiations.

37. In the *United Kingdom*, the sharing of identifiable personal health data is permitted among public authorities and the Information Commissioner, who is responsible for the UK Data Protection Act, advises public agencies on data sharing. There is a new initiative to create a National Health Service Information Centre for Health and Social Care where identifiable personal health data among several public authorities, the NHS, the Office for National Statistics and the Cancer Registry has been shared for the purpose of facilitating approved data linkage studies.

38. In the *United States*, federal authorities may enter into agreement with one another to share identifiable data. These agreements must conform to the legislative requirements of each of the participating authorities. The National Centre for Health Statistics negotiated an agreement for the sharing data with the Centre for Medicare and Medicaid Services and the Social Security Administration for the purpose of a data linkage study. The negotiated agreement took two years to complete.

39. In *Canada*, provincial data custodians can enter into agreement with federal agencies for the sharing of identifiable personal health data to build national databases. The Canadian Institute for Health Information is able to build national identifiable personal health databases by entering into agreements with each of the provincial government authorities for the sharing of identifiable data. While not bound by provincial laws, CIHI complies with provincial data protection and legal requirements in order to negotiate these agreements. Negotiations can take years to complete.

40. In *Singapore*, the law permits public authorities to share identifiable personal health data with another public authority. The authorities involved would enter into an agreement. To date, the Ministry of Health has not shared identifiable data with another government authority.

41. *Italy* has 19 territories and 2 provinces, each with local health authorities that process personal health data for their area. It is very difficult to engage in research with regional data because it is difficult to know how to approach the region with a proposal and what their requirements are for approval. The lack of adequate mechanisms makes it almost an impossible task, even for official institutions, to share data and information across multiple regions.

42. In *Cyprus*,[4] it is currently difficult for the Ministry of Health to obtain data from another ministry. The Health Ministry has been able to gain access to the population registry, which is in the custody of another ministry, but this access has been on a project-by-project basis. The barrier is that ministries

[4] See footnotes 1 and 2.

understand that the law does not permit the sharing of personal data. Specifically, the European Directive has been interpreted as not permitting sharing among ministries of identifiable personal data.

43. In *Finland*, for some national data collections, health authorities and physicians are required by law to collect the data and to provide it to the government. In practice, however, the Finland National Institute for Health and Welfare works actively to engage service providers in this collection effort by meeting with them and developing information to demonstrate to them why the information is important and how it will be used. This includes generating analysis at the local or regional level. Finland reports that it was difficult to establish disease registries at first. For example, it took ten years of negotiation to reach agreement with service providers to establish the first medical birth registry 25 years ago.

44. In *Denmark*, the law permits the sharing of identifiable personal data and the National Board of Health has shared data with Statistics Denmark for the purpose of specific projects requiring data linkages.

National infrastructure for data linkage and analysis

45. Record linkage involves linking two or more databases using information that identifies the same patient or the same person. 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. A specific type of record linkage, often referred to as deterministic linkage or exact matching, involves using a unique identifier or set of identifiers to merge two or more sources of data. In health linkages, the identifier used is often a unique patient identifying number or UPI. When a unique patient identifying number is consistently applied and recorded with few errors, this type of record linkage yields the highest quality and most accurate results.

46. Thirteen countries reported a national number that uniquely identifies patients. In nine countries, the number is used for health care encounters and other governmental purposes, such as social security and taxation. The *United States* reports the Social Security Number as a unique identifying number that can distinguish patients in public health-care programmes such as Medicare and Medicaid. The SSN, however, is not used generally for health-care encounters in the United States and is therefore not a national identifying number for health care services. In three countries, *Canada, Portugal* and the *United Kingdom*, the identifying numbers are exclusive to the provision of health services and are not used for taxation and social security. In *Canada*, the provincial HIN will change when the individuals move province and there is no linkage of old to new HIN numbers across provinces. As a result, record linkage studies that depend on the health insurance number experience a bias resulting from inter-provincial mobility. U.K. respondents to the telephone interview for this study were not sure if the NHS number issued to U.K. residents is a unique number that would be maintained when an individual moved within the U.K. or if it would change if an individual moved country, producing a similar bias to that experienced in Canada.

Table 1. National number that uniquely identifies patients and the main uses of this number

	Name of the unique identifying number	Main uses of the identifying number
Belgium	INSZ NISS	INSZ NISS is a national person identifier (national number) used for various purposes, such as health care, social security, and tax.
Cyprus*	Civil Identity Card Number	The Civil Identity Card Number is used by almost all government departments for administrative purposes, including the Ministry of Health, tax and social security.
Canada	Health Card Number	The provinces and territories assign a health card number that is a unique patient number for all publicly funded health-care encounters. There is also a unique Social Insurance Number assigned nationally for tax and social security purposes that is not used for health care.
Denmark	CPR NR (Central Person Register Number)	Used for 'everything' in relation to national and local governments including health care. Also banks and other business identifications etc.
Finland	Personal Identity Code	The personal identity code is used in practically all data collections in public services, such as health care, social welfare services, education, justice etc.
Italy	TS number	TS number contains both a health number and a tax file number and has nearly universal coverage of the population. It is managed through a publicly

		owned private company, SOGEI that could be considered as a trusted third party.
Korea	Resident Registration Number	Resident Registration Number (RRN) is assigned to each individual upon his/her birth and contains various information including birth date, gender and location of birth. RRN is used in virtually all aspects of life, including economic activities, for personal identification in various documents and communications in Korea.
Malta	Identification Number ID No	ID No is a unique identification number used throughout the country for all purposes including electoral lists, taxation, social security, etc. It is based on the registration number at the Public Registry.
Portugal	Número de Utente do Servico Nacional de Saúde	This number is used throughout the country for access to national health service care and benefits.
Singapore	National Registration Identity Care Number (NRIC)	NRIC is used for identification, government procedures, and some commercial transactions (e.g. the opening of a bank account)
Sweden	Personnummer (Personal Identity Number)	Personnummer is the main identifier used for all official purposes in Sweden (tax, social welfare, health care, living conditions, education and so on)
United Kingdom	NHS number Scotland also has the Community health index (CHI) number	Everyone registered with the National Health Service in England, Scotland and Wales is issued a unique NHS number. The NHS number is not used for tax/social security purposes. In Scotland, the CHI system was set up for administrative purposes to track patients registering with GPs.
United States	Social Security Number	The SSN is issued to U.S. citizens, permanent residents, and temporary (working) residents and its main purpose is for taxation.

Source: OCED HCQI Questionnaire, Secondary Use of Health Data, September and October 2011 and, for Italy, follow-up telephone interview, October 2011

Note: *See footnotes 1 and 2.

47. Twelve countries reported a unique identifying number for patients exists currently within their national hospitalization databases and that this number could potentially be used for data linkage (Table C4). Eleven countries reported the same conditions for their cancer registry; and ten for their primary care, mortality and population census or registry data. Nine reported the same conditions for their mental hospital in-patient data and their population health survey data. Eight had this condition for formal long-term care data and seven for prescription medicines data. Only one country, however, had a unique identifying number that could be used for data linkage of patient experiences data.

48. There are new developments in three countries that have not been able to use a unique identifying number for record linkages, *Switzerland*, *Germany* and *Japan*. The current process in Switzerland involves the health care providers in the Swiss Cantons, who have access to patient names, dates of birth and sex, to create an encrypted identifier that cannot be reversed to reveal the identity of a person. The same algorithm is applied throughout the country and through time and is provided to the Federal Statistical Office (FSO) who uses it to enable data linkages. The algorithm has limitations. In particular, it does not account for name change, which creates a systematic bias in the data, particularly for women, where changes in marital status may result in name changes. There is a unique Social Security Number (SSN) in Switzerland that could potentially be used for data linkage in the future in an encrypted form. Recently, the Swiss Federal Statistical Office (FSO) sought an option of the Swiss national Office of Data Protection to determine if the FSO had the legal authority to process data using the SSN. The determination was that this use is in compliance with the health insurance law and could be in compliance with the law authorizing the FSO, if the FSO amends the ordinance that accompanies its authorizing legislation that specifies the data that the FSO is collecting. The FSO is pursuing this change in its ordinance. In *Japan*, there is a current proposal to introduce a uniform identifying number for tax and social security purposes, including health care. In *Germany*, there was an announcement by the government in October 2011 of the intention to introduce a unique health insurance number.

49. Other variables in a database can also be used to link records through a process of exact matching or through probabilistic matching. For probabilistic matching, 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 that the records match. A combined score represents the probability that the records refer to the same individuals. 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.

50. More countries reported having a set of identifying variables within their databases that could be used for record linkage than reported having a unique patient identifying number (Table C5). These variables included names, dates of birth, addresses or postal codes, sex, and dates of events. Not all of these identifying variables are available on all of the data, but all of the data have at least some of these identifiers. Thirteen countries reported having a set of identifying variables within their hospitalization databases and twelve reported these variables within their cancer registries and mortality databases. Eleven reported these are part of their population census or registry and ten countries reported these as part of their mental hospital in-patient data. Nine countries reported these identifiers within primary care data and eight reported these within formal long-term care data. Seven countries reported these identifiers within prescription medicines data and population health survey data. Only 2 reported such identifiers within patient experiences data.

Sub-national infrastructure for data linkage projects

51. In some countries, data linkage is commonly undertaken at the level of regions, states or within specific networks of health-care organizations. Networks of health-care organizations, such as the U.S.A. health-care organization network Kaiser Permanente, offer a broad range of health-care services and can conduct research where patient data is linked across the different health care facilities they operate.

52. Seven countries reported sub-national data linkage activity at the state or region level (Table C9). *Canada* reported regular health-related data linkage activity across all the major types of health data in nine of the ten Canadian provinces and involving a unique patient identifying number, the provincial Health Information Number. Canada also reported that these provinces have a broader range of projects using data linkage because the provinces have access to more detailed and comprehensive data than is available nationally.

53. *Germany* reported data linkage project activity at the state level involving cancer registry, mortality, population health survey and other data. Examples include projects related to the development of a mortality index in Bremen state; sickness fund data linkages in Hessen; and linkages involving population health surveys in Augsburg and Essen. The states of Bremen and Hessen are undertaking health-related data linkage studies on a regular basis. These state-level linkages benefit from unique patient identifying numbers. For both legal and administrative reasons, states are able to undertake linkages while there is no activity at the national level. *Portugal* and *Japan* reported sub-national infrastructure for data linkages within cancer registries.

54. *Sweden* also reported data linkage activity within some of the 21 county councils, such as the Skåne Region and the West Region and that these regions are able to undertake a broader range of data linkage activities than can be undertaken at a national level. For example, the West Region has a primary care register that may be linked.

55. The *United States* reports that each state (plus DC) has a wide variety of data users, data sources and products and may well be undertaking data linkage projects. Further, states have Social Security Numbers that might be used to facilitate linkages along with Medicaid identifiers. Whether or not the states are undertaking a broader range of data linkage activities than are taking place at the national level cannot be determined without an extensive survey. However, the medical and health services literature shows a wide variety of research studies by government, academia, health care quality organizations and industry in the United States.

56. The *United Kingdom* also reports sub-national data linkage activity in the region of Tayside Scotland. This local area does not, however, have a broader range of data linkage projects than are possible at the national level in Scotland.

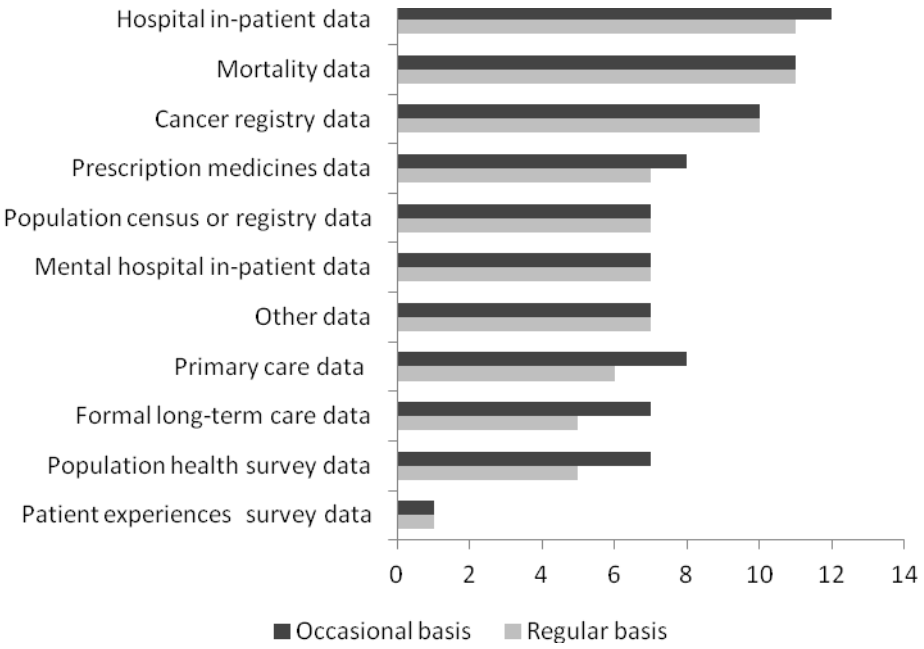
57. Six countries, *Belgium, Canada, Germany, Portugal, Singapore and the United States* reported networks of health-care organizations conducting data linkage projects with their own data (Table C10). Belgium reported this activity within networks of hospitals. Germany reported this activity for several statutory health insurance funds such as Barmer-GEK, AOK and the Bremen Institute for Prevention Research and Social Medicine, BIPS. Portugal reported this activity within Integrated Delivery Services. The United States reported this activity among large health-care insurers including Kaiser-Permanente, Puget Sound, Havard Health Plan and others. *Singapore* reported that public health-care providers undertake this type of work on an ad hoc basis.

Data linkages for public health research and health-care quality monitoring

58. Most countries with variables within their national databases that would permit data linkages have conducted data linkage projects. Overall, most countries reported regular or occasional record linkage projects involving hospital in-patient data, mortality data and cancer registry data (Tables C6 and C7). Half of the countries also reported record linkage studies with all other major types of data, with the exception of patient experience surveys where data linkage has occurred in only one country.

59. Ten countries report undertaking national data linkage projects to monitor health care quality involving hospital in-patient data and cancer registry data and eight countries include also the linkage of mortality data for this purpose (Table C8). Using data linkage techniques to monitor health care quality in other areas of health care is much less common with five countries monitoring pharmaceutical care quality; four monitoring mental health care quality; and three monitoring the quality of primary health care and formal long-term care. Three report linkages to population health surveys or population census/registry data to support health care quality monitoring and only one has benefited from linkages to patient experiences data for this purpose.

Figure 2. Number of countries reporting national data used to conduct record-linkage projects on an occasional and on a regular basis



Source: OCED HCQI Questionnaire, Secondary Use of Health Data, September and October 2011

60. Seven countries have a regular occurrence of data linkage projects involving many national databases (Canada, Denmark, Finland, Republic of Korea, Singapore, Sweden, and the United Kingdom). In these countries, a unique patient identifying number is available to facilitate the linkages (Table 1). The United States and Switzerland also have a regular occurrence of data linkage projects with a number of databases and rely on sets of patient identifying information to establish links. Belgium (4 databases), Cyprus[5] (5 databases) and Portugal (4 databases) have national databases with patient identifying numbers and/or other patient identifiers, but engage in data linkage on a regular basis with only two of the available databases. Japan has many databases with a unique identifying number that could be used to establish linkages but does not undertake data linkages on a regular basis. Germany has a few national databases with a unique identifying number that could be used for data linkages and does not undertake national data linkage projects on a regular basis.

61. *Finland* reports that hospital in-patient data is linked to formal long-term care data on a regular basis to get complete information on institutionalised care; cancer registry data is combined with mortality data to complete the data with all cancer cases; and data on deaths is combined with the Medical Birth Register and the Register on Congenital Malformations to get more exact information on perinatal and infant deaths. To monitor health care quality, examples include combining registers to get information on the consequences of the use of medicines during pregnancy on the health of newborns; to benchmark hospital health-care quality performance for major diseases and medical conditions, such as stroke and very premature births (see chapter 5); and to monitor life-expectancy among patients with severe mental health disorders who have been hospitalized. This last project was a multi-country study with other Scandinavian countries (see section 4.5. multi-country studies).

[5] See footnotes 1 and 2.

Table 2. Distribution of the regular occurrence of health-related record linkage projects by availability of databases with patient identifiers

	Most national data with a unique patient identifying number (UPI)	Most national data with other patient identifiers	Some national data with a unique patient identifying number (UPI)	Few national databases with patient identifiers
National Data linkage projects on a regular basis with 7+ national databases	Finland, Republic of Korea, Sweden, United Kingdom	United States		
National Data linkage projects on a regular basis with 5-6 national databases	Canada, Denmark, Singapore	Switzerland		
National Data linkage projects on a regular basis with 3-4 national databases	Malta			
National Data linkage projects on a regular basis with 2 national databases			Belgium, Cyprus*, Portugal	
No Data linkage projects on a regular basis	Japan			Germany

Source: OCED HCQI Questionnaire, Secondary Use of Health Data, September and October 2011

Note: *See footnotes 1 and 2.

62. The *Republic of Korea* reports an extensive programme of regular health-care quality monitoring using data linkages. Indicators from the linkage of hospital in-patient data to mortality data include thirty-day case fatality for acute myocardial infarction and thirty-day post-operative mortality for major types of surgery. Linkages of mental hospital in-patient data to hospital in-patient data enable monitoring hospital re-admissions for mental-health patients; and further linkage to prescription medicines data enable monitoring health outcomes of prescribing to mental-health patients. Outcomes of prescribing patterns in primary care are monitored through linkage of prescription medicines and primary care databases. Korea also links the cancer registry data to mortality data to assess the relative survival of cancer patients and links long-term care data to survey data on the activities of daily living to estimate the percentage of patients with reduced activities of daily living (see chapter 5).

63. *Sweden* also reports a comprehensive programme of data linkages that facilitate health care quality monitoring including regular linkages of all registers to mortality data; linkages of patient registry data to the prescribed drug register; and the cancer register to the patient register (see chapter 5). *Denmark* reports a similar data linkage capacity including linkages to more than 50 national clinical quality databases.

64. The *United Kingdom* has the most comprehensive suite of national data among the countries that participated in this study; however, the coverage of these databases is often limited to one or two of the member countries. In Scotland, hospital in-patient data, cancer data, mental hospital in-patient data and mortality data are maintained as a permanently linked database. Prescription data has only recently become available at record level with a UPI in Scotland and will now be regularly linked. Population health survey data is used regularly in research linkages in Scotland. Scotland reports using linkage to monitor outcomes of health care including HEAT targets, such as monitoring readmissions and deaths among coronary heart disease patients. In England, hospital data is linked to mortality data on a monthly basis. England monitors hospital standardized mortality ratios that will be replaced, in future, with a summary hospital-level mortality indicator (SHMI). Cancer incidence data in England is routinely linked to mortality, hospital treatment (surgery and radiotherapy) and, for a proportion of the population, to primary care data. Birth

notifications are linked to birth registrations (e.g. to determine prematurity) and to death registrations in England and the cancer registry is linked to mortality data. England produces a thirty-day post-operative mortality rates for patients following colorectal cancer surgery. In England and Wales, the ONS Longitudinal Study (LS) has linked a 1% sample of the population census in 1971, 1981, 1991 and 2001 across censuses and to births, deaths and cancer registrations. The study can be used to understand the distribution of health outcomes by census population characteristics as well as changes in characteristics and health outcomes over time. Wales has linked births to hospital delivery records (see chapter 5); and the cancer registry to mortality data. The linkage of hospital in-patient data to other databases is under development.

65. *Canada* also has a number of national databases that are regularly linked using a unique health care identifying number administered by each province. Hospital in-patient data are often linked to other types of health care including emergency room visits (see chapter 5); and population health surveys are routinely linked to in-patient hospitalization data and to mortality data. At the provincial level, data linkage activity to inform about population health and health-care quality is extensive (see chapter 5).

66. The *United States* reports the regular creation of files linking hospital records, the cancer registry and the population census to mortality data; and population health survey records to mortality data and to health care records for Medicare and Medicaid enrolees (see chapter 5). National health care quality monitoring from data linkages includes cancer survival rates, thirty-day mortality following in-patient hospitalizations, and infant mortality.

67. *Switzerland* reports the linkage of hospital in-patient data, mental hospital in-patient data, formal long-term care data, mortality data and the population census (see chapter 5).

68. In *Belgium*, hospital data is regularly linked to hospital expenditure data; and cancer registry data is linked to mortality data, to health insurance nomenclature, to hospital in-patient data and to cancer screening. Databases on cystic fibrosis and neuromuscular disease patients are linked to the population register to capture year of birth, district, sex and deaths. Belgium reports data linkages to produce process and outcome indicators for breast, testicular, and rectum cancers with on-going work on oesophagus and stomach cancers (see chapter 5). Linkage has also been used to assess GP performance.

69. In *Cyprus*,[6] from 2004 on-ward mortality data have been regularly linked to the Cancer Registry in order to obtain the follow up data necessary for cancer survival estimation. Survival calculations, however, have not yet been produced.

PROTECTION OF PRIVACY IN THE COLLECTION AND USE OF PERSONAL HEALTH DATA INCLUDING DATA LINKAGES

70. All countries report legal and policy restrictions on the collection and use of personal health data that reflect the importance of the protection of data privacy and confidentiality. This section is not in any way exhaustive of the full legal frameworks in place within countries. What it presents, instead, are the views of officials responsible for data protection and health researchers that were developed from their

[6] See footnotes 1 and 2.

personal experience of working within their legal frameworks to make decisions about or to undertake projects requiring the secondary use of personal health data.

Guiding principles and legislations

71. All countries report a legislative environment with specific pieces of legislation that relate to the protection of personal information in general and, for some, additional legislation specific to health data protection. New legislations and privacy policies have all been influenced by the 1980 publication of the OECD privacy guidelines and these guidelines are still recognized 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 the following eight guiding principles (Box 2):

1. Collection Limitation
2. Data quality
3. Purpose specification
4. Use limitation
5. Security safeguards
6. Openness
7. Individual participation
8. Accountability

72. These principles were subsequently reflected in the 1995 data protection directive of the European Union 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.

73. 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.

Box 2: 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 unauthorized 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 him; 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 him 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, *Policies for Information Security and Privacy*, 2009

74. Among the non-European countries in this study, the United States reports a federal *Privacy Act* with data protection requirements for federally-held personal data and the *Health Insurance Portability and Accountability Act (HIPAA)* which specifies data protection requirements for personal health data in the United States. Canada reports a federal *Privacy Act* with data protection requirements for personal data and a federal *Personal Information and Protection of Electronic Documents Act (PIPEDA)* with specific data protection requirements for the transborder movement of personal data linked to a commercial enterprise.

75. Japan reports a *Privacy Protection Act* that governs the protection of personal information. The Republic of Korea has a new *Personal Information Protection Act* that specifies the requirements for the protection of personal data. Although this *Act* has not yet been enacted, its provisions have already taken

effect. In Singapore, the common law regulates the collection and use of identifiable personal health data. This is complemented with the *National Disease Registry Act* which specifies the data protection requirements related to disease registry data; and the *Computer Misuse Act* which requires public and private entities with electronic data to protect data confidentiality and security.

76. Sub-national legislations related to the protection of personal information or personal health information are reported for the states within the United States and among the Canadian provinces.

77. In addition to national data protection offices, Canada reports privacy commissioners within the 10 Canadian provinces; Switzerland reports an Office of Data Protection within each of the 26 Cantons; and Germany reports data protection authorities in each of the 16 states.

78. Most countries also report authorizing legislations that relate to the work of health ministries, statistical offices and other public authorities that also specify requirements related to data protection. Some have legislations at a much finer level as well, such as enabling legislation for a particular disease registry.

Privacy principles in practice

79. The collection and use of personal health information follows the principles of privacy protection in all of the participating countries. All countries, however, have areas where the application of privacy protections could be improved. Also, and importantly, some countries have applied privacy principles in a way that unnecessarily impedes privacy-respectful health research in the public interest. In these countries, reforms could likely facilitate greater public benefit from health information infrastructure without deterioration in public confidence.

80. Health monitoring and research often requires the use of health care databases originally collected for the purpose of the administration of the health system or for direct patient care. Health monitoring and research uses may not have been considered when the data were collected and persons from whom the information was gathered were not, consequently, informed. These databases often represent thousands or hundreds of thousands of patients and re-contact to ask a consent question is either impossible and/or financially infeasible for a country. These realities can place health research in question of disrespecting collection and use principles of data protection. These principles include that personal health data should be collected in a fair manner where individuals are aware that the information is being collected and are aware of the purposes of the information collection; that the subsequent use of the personal information should conform to the same purposes and not deviate from them; and that individual data subjects should provide consent to any new use of the personal health information or the new use should be authorized in law.

81. In *Belgium*, the Privacy Commission grants authority to collect and use identifiable personal information without consent. After the introduction of the European Directive on the Protection of Personal Health Data, the Privacy Commission advised the Cancer Registry that it could no longer process identifiable personal health data and that the only way it could continue normal operations would be to draft authorizing legislation and reapply for permission. The legislation authorizing the Cancer Registry clarifies that patient consent is not required to create the registry nor to link or analyse the registry data. The years when the legislation was being drafted, and normal operations were suspended, involved degradation in the quality of the registry, coupled with a resource-intensive process to try to maintain quality (see chapter 5).

82. In *Italy*, when the European Directive was first introduced, the possibility to conduct health research involving identifiable personal health data was reduced. Under the first Italian *Data Protection*

Act that came into force in 1997, personal health data should be de-identified; and only if it was impossible to do so, should identifiable health data be processed. In 2004, a *Data Protection Code* was introduced that included a chapter on the specific case of data processing in the health sector. This defined categories of the processing of identifiable personal health data that would be considered in the substantial public interest. This code permits the processing of identifiable personal health data if the data subject has given consent or if law authorizes the process. Currently, many Italian regions have legislation that authorises them to develop disease registries from health care data without consent and to use the data for research purposes. Further in 2011, the Privacy Guarantor, who is the data protection authority, gave a general authorization to enable regions to process identifiable and sensitive health data for research purposes. National birth and death registries exist in Italy but it remains very challenging to build national disease registries in Italy because a national registry that consolidates data from regional registries would be constructed from regional data that was collected without informed consent. As in Belgium, any national registry in Italy would require its own authorizing legislation to be approved by the data protection authority. While regions have been authorized to conduct research and analysis with registry data, there is a growing concern that the Privacy Guarantor may revoke this approval. This concern has put a chill on health research in Italy, as many regions are becoming reluctant to participate in research studies.

83. In *Germany*, each of the 16 states has a data protection authority that is independent from government. Thus, data protection requirements vary across these different jurisdictions. Each of the states also has its own legal framework that enables a cancer registry to be created at the state level without informed consent. There is no national cancer registry and any amalgamation of data from states for particular research projects requires each state's authority to proceed (see chapter 5).

84. In *Cyprus*,[7] the processing of personal health data and data linkages have been very restricted following the European Directive. EU regulations for countries to collect causes of death and to provide statistical results to Eurostat, however, imply that the collection of these statistics is lawful and the Ministry of Health has complied with the request. The preparation of death statistics requires the existence of a registry. While a death and a cancer registry involving the processing of identifiable personal health records currently exist at the national level in Cyprus[8], a draft law is being prepared that would provide a framework for the Ministry of Health to be able to continue to maintain a death and a cancer registry. This is taking place out of a concern that these registries could be at risk of being determined illegal. *Portugal* reports a similar situation where access to data across data custodians is limited and data linkages are not allowed.

85. In *Sweden*, *Denmark* and *Finland*, legislation enabling a range of health registries and social welfare registries is in place. This legislation makes participation in the registries mandatory and enables the identifiable data to be processed without informed consent. In *Sweden* and *Finland*, if a patient wants to have their personal data removed from a registry they may appeal to the national health authorities. Consent processes have changed with respect to health care quality registers in Sweden. In the past, patients were informed of the use of their data through information and brochures. Now, hospitals are asking patients for their consent to use their personal data for research or statistical purposes. In *Denmark*, a patient cannot request the removal of their data from patient registries. Patients may, however, ask that their contact information never be provided for research projects where they would be contacted to answer a survey. In *Finland*, the website of the NIHW is used to communicate with the public the data files that are prepared, where the data comes from and how the data is used.

[7] See footnotes 1 and 2.

[8] See footnotes 1 and 2.

86. In the *United Kingdom*, all data custodians must register their collections of personal data with the UK Information Commissioner, who is responsible for overseeing the *Data Protection Act*. Schedule 3 of the UK *Data Protection Act* lists conditions that may apply to justify the processing of personal health data without consent. These include that it is necessary for prevention, diagnosis, medical research, patient care or the management of the health care system. The Act outlines levels of consent requirements for circumstances where patient consent would not be required (such as the communication of information about communicable diseases to authorities) to circumstances where consent would always be required. The processing of personal health data falls between these two ends. Data linkages may be undertaken without obtaining patient consent when government has collected the datasets involved. The onus is on data custodians to communicate with the public about how their data is being used. For example, National Services Scotland (NSS) has information about data collection and use on their website including any privacy notices. The NSS is considering increasing this public communication to also include a description of when data projects are using identifiers or there are data linkages and who has been provided access to the data. Another example is within the Office of National Statistics, where posters are put up in birth and neo-natal units in health facilities and a leaflet is distributed to new parents that informs about the ONS, its mandate, the collection and the use of birth registration information including data linkages and some recent findings. Patients have the right to refuse to have their data used for health research. For example, the NHS Information Centre reports the rare occasion of a request to supress a patient's hospitalization records.

87. In the *United States*, there is no central national authority for granting of approval for uses of personal health data. *The Health Insurance Portability and Accountability Act (HIPAA)* applies to certain covered entities and governs when patient consent is required and when personal health data may be used without consent. Covered entities include health plans, health care clearinghouses and health care providers who electronically transmit health information in connection with transactions including billing and payment for services or insurance coverage. Under *HIPPA*, written consent of data subjects is required to use or to disclose identifiable personal health information unless this use is for public health purposes or the data custodian's internal review board or a privacy board has approved it (National Institutes of Health, 2011). For example, Kaiser Permanente is a covered entity and it meets *HIPPA* requirements for written consent by including the collection and use of personal information within the terms of the membership agreement that is signed by individuals joining the Kaiser health insurance plan (see chapter 5).

88. The U.S. National Centre for Health Statistics has its own authorizing legislation that permits the collection and use of personal health data. NCHS has an Internal Review Board (IRB) that approves data collections and data linkage. For any linkages that would involve health care administrative records, such as records from the Centre for Medicare and Medicaid Services, *HIPPA* requires that the linkage must conform to the terms of the statement signed by enrolees in these insurance programmes. This statement described the uses of the data and represents the patient's written consent. For example, the linkage of health care administrative files and immigration files may be determined to be outside of the terms of the signed statement and therefore not permitted. For surveys, the NCHS administers a question to respondents asking for their consent to link their survey responses to other health care and vital event databases for statistical purposes. In the past, survey respondents were asked for their Social Security Number but were not asked for permission to link their survey responses to other health data. Typically, only 50% of respondents would provide the number. Those individuals were assumed to have consented to data linkage and were the only records eligible for linkage. After the NCHS changed the process and began to ask survey respondents if they would consent to data linkage, the proportion of respondents saying yes to the National Health Interview Survey grew to just under 90%. Further, for records where respondents have consented and the Social Security Number is missing or incorrect, probabilistic matches are now possible.

89. In the *Republic of Korea*, personal health data may be collected and used with patient consent and, where authorized by law, without patient consent. Under the *National Health Insurance Act* and the

Cancer Control Act, personal health data is authorized to be processed without consent by public authorities. In *Singapore*, patient consent is required for uses of administrative data that are beyond direct patient care. For public policy purposes, this requirement is met by informing patients. For example, when a patient is admitted to hospital they are informed about the uses of their data. There is also information provided to patients making a claim under the national health insurance programme.

90. In *Canada*, the *Personal Information and Protection of Electronic Documents Act (PIPEDA)* was introduced in 2000. *PIPEDA* governs the sharing of electronic health records across jurisdictional boundaries when those records originate from a commercial source, which can include health-care providers. Organizations covered by this federal *Act* must obtain the consent of individuals when they collect, use or disclose personal information unless they are authorized to do so by another law. The introduction of the *Act* created ambiguity as to the legality of health research activities involving administrative health data without the express consent of data subjects. The Federal Privacy Commissioner's office eventually made a determination that there could be secondary use of personal health information without patient consent in situations where the use of the data could be demonstrated to be in the public interest.

91. Some provinces have since introduced legislation governing the protection of personal health information. For provincial laws to supersede the federal *PIPEDA*, a prerequisite is that they must be similar in spirit. Ontario was the first province to introduce its own legislation, the *Personal Health Information Protection Act (PHIPA)*, in 2004. *PHIPA* clarifies in law that certain prescribed entities are able to collect and use personal health data without patient consent. The Institute for Clinical and Evaluative Sciences (ICES) at the University of Toronto, for example, is a prescribed entity and receives identifiable data for research purposes from a variety of public authorities in Ontario and receives Ontario-specific identifiable data from Statistics Canada (see chapter 5). ICES is authorized to process the data and conducts research and publishes research based on data linkages. To receive data from the province of Ontario without patient consent, the Canadian Institute for Health Information (CIHI) also needed to become a prescribed entity under *PHIPA*. As other provinces introduce similar legislation, CIHI works with the provinces to ensure that the legislation will permit CIHI to continue to receive transfers of personal health data to build and use national databases for statistical and research purposes. The province of Quebec has legislation that does not permit the transmission of identifiable personal health data to third parties, which includes CIHI. CIHI has appealed to Quebec for a resolution.

Data linkage activities and compliance with legislation

92. All countries have entities with the legal authority to conduct record linkages for public health and health services monitoring and research under certain restrictions that relate to legislative requirements for data protection.

93. In *Finland*, the National Institute of Health and Welfare (NIHW) and Statistics Finland are both authorized by the data protection authority to conduct data linkages using identification numbers. In practice, the NIHW receives identifiable data from the statistical office and conducts data linkages. National identifying numbers are used in initial processing of the data to edit the data and check the data for errors. When the data is clean, the identity numbers are encrypted and the encrypted numbers are used to perform linkages for approved projects. There was a case, however, where a project involved the linkage of criminal data to health data. For this project, the data protection authority required the linkage to be undertaken by a third party. While exceptional, in Finland it is possible for a request from an external researcher for access to identifiable health data to be granted.

94. In *Sweden*, the National Board of Health and Welfare conducts data linkages using identification numbers. Analysts within government and external researchers with approved projects are only provided

access to de-identified data. Similarly, in *Denmark*, the National Board of Health conducts data linkages. In cases where databases of the National Board of Health would be linked with databases from Statistics Denmark, identifiable data would be provided from the board to Statistics Denmark who would conduct the linkage and de-identify the data. Only de-identified data is provided to researchers within and outside of government.

95. In the *United Kingdom*, national data custodians most often undertake data linkages involving national data. However, UK law does not rule out the possibility that a non-governmental researcher could receive approval for access to identifiable data and conduct a data linage. Linkages most often take place using the unique National Health Service number or, in Scotland, the unique Community Health Index number. Probabilistic linkage is used where deterministic linkages fail or when unique numbers are missing. In England and Wales, a new center has been created to facilitate health-related data linkages, the NHS Information Centre for Health and Social Care. This Information Centre is a government initiative to facilitate research with personal health data that are in the public's interest. The Centre receives identifiable data that has been shared from multiple public entities including the NHS, the ONS and the Cancer Registry. Projects that have been approved by the U.K. Information and Governance Board may have their linkages undertaken by this centre. These projects may be on behalf of public or private researchers. Only de-identified data is provided to clients for research.

96. In *Belgium*, as a result of the legislation specific to the cancer registry, the Privacy Commission has approved the cancer registry to collect identifiable personal health data and to link the data and then to conduct analysis of de-identified data. In general, however, data linkage takes place within the E-health Platform which is a third party authorized by law to access and use identifiable health data and who is trusted to undertake data linkages that are approved by the Privacy Commission. Only de-identified data is provided to governmental and non-governmental researchers for analysis.

97. Data custodians undertake data linkages in the *United States*. The NCHS conducts data linkages among its own databases with the approval of its Internal Review Board (IRB). There is no unique patient identifying number in the United States; however, it is sometimes possible to conduct linkages using Social Security Numbers. Linkages are typically probabilistic linkages that depend on a set of identifiers in the data (names, dates of birth, marital status, place of birth and race). For a linkage of NCHS survey respondents to health care administrative data held by the Center for Medicare and Medicaid Services (CMS), the linkage was conducted in steps involving three governmental organisations using a deterministic linkage method. Records of respondents in the survey who consented to data linkage were shared with the Social Security Administration; who linked the data to the social security database and corrected any errors in the Social Security Numbers captured on the survey. The corrected data was then sent to the CMS who conducted a deterministic linkage to Medicare and Medicaid records and then removed the Social Security Numbers from the linked file and provided the linked file back to the NCHS. Only de-identified data is ever provided for research and the de-identification process is very strict (see below).

98. In *Canada*, CIHI undertakes data linkages at the national level involving health-care administrative data. The main linkage key used is the provincial Health Insurance Number. Health Insurance Numbers are encrypted during data processing at CIHI and deterministic linkages are undertaken using these encrypted numbers and other identifiers such as birth dates and dates of treatment. Only de-identified data is ever provided to internal data analysts or external researchers. In cases where linkages would require the databases of Statistics Canada, identifiable data has been shared with Statistics Canada who has undertaken the linkage. In some cases, such as for mortality and cancer registry data, linkages are primarily probabilistic due to the inavailability of health insurance numbers (Statistics Canada, 2006). Such data sharing arrangements with Statistics Canada only take place through negotiated agreements and with the approval of the provinces whose data would be involved (Statistics Canada, 2010). Only de-

identified data is ever provided for research and the de-identification process is very strict and similar to the U.S. NCHS (see section 4.2.4).

99. In *Singapore*, there is no national authority for the conduct of data linkages and different governmental institutions are undertaking data linkages including the Ministry of Health. In the ministry, effort has been made to automate data linkages to as high a degree as possible through deterministic matching using the National Registration Identity Care Number (NRIC). The computer algorithm will also automatically de-identify the data and produce a file ready for analysis. There are a small number of records that cannot be automatically linked as sometimes NRIC numbers are missing, such as among individuals who have not completed the citizenship process. For these cases, probabilistic techniques are used to link the records. Only de-identified data is every provided to internal analysts or external researchers and typically only under controlled conditions (see below).

100. In *Japan* there are no reported legislative barriers to undertaking data linkages and the National Institute of Public Health reports linkage is technically possible involving hospital, pharmaceutical, primary care data and population survey data. The Ministry of Health, Labour and Welfare in Japan, however, reports removing all identifiers from health care databases and rendering record linkage impossible.

101. In *Cyprus*,[9] data custodians are authorized by law to undertake data linkage and the Ministry of Health has been able to link the Cancer Registry to mortality statistics to create the possibility to generate information about cancer survival. The Civil Identity Number is the principle key used to conduct the linkages. As was noted earlier, a restriction on sharing of data across government authorities is the main barrier to data linkage.

102. In *Germany*, data linkages take place at the state level and not at the national level and only when authorized by state law. For example, each state has legislation that enables cancer registries. In German states, names, addresses, dates of birth and place of birth are used to establish linkages probabilistically. The same national peudonomization algorithm is used by all of the German states to render names anonymous. Thus, with approval, it is possible to share de-identified records across state lines and correct for the bias in the registries that would otherwise occur from patient mobility. Only de-identified data is provided to researchers within and outside of government.

De-identification of data

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

104. In *Finland*, data is considered de-identified when the identity number has been encrypted and names have been removed. Researchers outside of the National Institute of Health and Welfare (NIHW) with approved projects receive data with encrypted identity numbers to conduct their analysis. In *Sweden*, data is de-identified by the National Board of Health and Welfare by removing national identity numbers, names, addresses and full dates of birth. Files provided to analysts within government and outside of government contain a study number that has been assigned in place of the identity number as well as some personal information on sex, age and home community. In *Denmark*, the National Board of Health data is de-identified by removing names and exact addresses. The national Central Person Register number, however, will remain on the analytical file. This number reveals the sex and birthdate of the person.

[9] See footnotes 1 and 2.

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

106. In *Canada*, the Canadian Institute for Health Information (CIHI) provides the algorithm to provinces to encrypt the health insurance number within health care administrative data before the data is transferred to CIHI. Some provinces provide health care administrative data with original health insurance numbers included. In this case, CIHI will encrypt the health insurance number using the same algorithm. For approved projects, researchers will receive a file without names or exact address and encrypted health insurance numbers. Researchers are able to conduct the linkage of files with the encrypted numbers themselves.

107. In the *United States*, the National Centre for Health Statistics considers that data is de-identified when the risk of potentially re-identifying persons within the data has been reduced. This includes removal of identifiers, such as names, exact addresses, full dates and any identifying numbers and also a careful review of possible combinations of remaining sensitive variables within the data file that may indirectly lead to the disclosure of the identity of a person. Individual-level data that has been de-identified to this standard can be made publicly available and can be disseminated over the Internet to the public. For example, the linkage of population survey data to death data has been released as a public-use micro data file. Often, however, the level of detail that is required for an approved research project would create a re-identification risk that is too high for the NCHS to release the data to the researcher. Instead, the NCHS has created a network of secure research data centres that researchers with approved projects must use. Similarly, in *Singapore*, the Ministry of Health encourages researchers to apply for access to de-identified data within the Ministry's secure data lab (see below).

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

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

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

110. In *Canada*, Statistics Canada also maintains a network of secure Research Data Centres across the Canadian provinces with similar features to the U.S. RDCs (Statistics Canada, 2011). Researchers with approved projects may only have access to de-identified data with a high re-identification risk within the

RDCs. Canada does not yet have a remote data access option, but is beginning to pilot options that may enable this type of access in the future.

111. The *Singapore* Ministry of Health has also established a secure data laboratory that has been available for the past year. The ministry was particularly concerned with the re-identification risk resulting from fulfilling data linkages for hospital-based researchers to data that had been collected by the researchers. All approved research by government and non-government researchers involving access to personal health data must take place within the lab. Only aggregated results that have been vetted by a ministry staff member may exit the secure lab. A researcher may apply to the ministry for approval to have access to de-identified data outside of the lab. This request would require the approval of the ministry's internal review board and there would have to be a strong justification for granting the request.

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

Project approval process for data linkages

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

114. In *Singapore*, the Ministry of Health has an Internal Review Board that reviews proposals and grants approval for projects internal to the ministry and those from other government and non-government researchers. Approval is typically granted on the basis that access to the data will take place within the ministry's secure data lab.

115. In the *Republic of Korea*, the Ministry of Public Information and Security approves data linkage projects proposed by Health Insurance Review and Assessment Service (HIRA) on a project-by-project basis. Government and non-government researchers external to HIRA may apply to HIRA for access to de-identified personal health data including linked data that HIRA has in its custody.

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

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

118. In the *United Kingdom*, the UK Data Protection Act provides the legal framework wherein a National Information and Governance Board (NIGB) was created to provide a national decision-making body on any projects undertaken in the public sector or on the private sector where the consent of the data subjects was not obtained and where the use of the data is not authorized in law. The NIGB Ethics and Confidentiality Committee acts as a national research ethics approval body for all data custodians responsible for health and social care data. Thus, projects initiated by the public or private sector can be reviewed for their conformity with the law and the relative balance between research that is in the public's interest and the respect of privacy principles can be weighed. For data files outside of the domain of health and social care, or for regions outside of NIGB jurisdiction (Scotland), the Caldicott Guardian would act as the approval body. Each custodian of personal data is required by law to have a Caldicott Guardian which is a senior official entrusted to protect data privacy and who is responsible for evaluating and approving projects requiring access to and use of personal data.

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

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

121. In *Cyprus*,[10] de-identified data, whether linked or unlinked, is not shared with researchers external to the Ministry of Health. This is because it is not clear which public authority would be able to approve the request. The Ministry of Health has asked the Offices of the Personal Data Commissioner and the response has been that the Commissioner does not grant permission but only acts as a control.

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

[10] See footnotes 1 and 2.

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

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

124. Virtually all countries that will provide researchers with access to linked data will consider such a request for approval. In all cases, however, the requesting researcher must be able to demonstrate that they had collected the data with the informed consent of the data subjects or had legal authorization.

125. Few countries make any exceptions. In the *United Kingdom*, a request where the researcher did not have informed consent could be reviewed for a decision by the National Information and Governance Board. In the *Republic of Korea*, any request where the researcher's cohort was obtained without consent would have to be reviewed by the Public Information and Security Ministry for approval. In *Belgium*, the Privacy Commission renders a decision on all project proposals and would hear the proposal.

126. In *Canada*, however, a researcher requesting a linkage of a cohort of data with health care administrative data would have to apply to each individual province to fulfil this request and could not be approved for linkage to the national databases of the Canadian Institute for Health Information. Linkages of a researcher's cohort to mortality and cancer registries, however, could be granted at a national level by Statistics Canada (Statistics Canada, 2006). In *Italy*, there are no routine or standardized procedures for a researcher to request a linkage of their own cohort of data to governmental databases and it seems that this type of project is impossible.

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

Data security within public authorities

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

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

    a. The Belgian Cancer Registry, the Cyprus[11] Ministry of Health; the Canadian Institute for Health Information; and the U.K. NHS NSS Scotland reported a requirement for new employees to sign a document that they will protect data confidentiality.
    b. The United States NCHS and the U.K. NHS Scotland reported an annual requirement for all employees to sign a document that they will protect data confidentiality.

[11] See footnotes 1 and 2.

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

    c. The U.S. NCHS has a staff manual on data confidentiality protection requirements
    d. Data security and privacy guidelines are communicated to all employees of the Republic of Korea HIRA using the internal network homepage.
    e. At the U.K. NHS NSS Scotland, standards for data protection and security are described in the document that employees must sign annually.

3. Levels of approved access to data for staff

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

4. Restricting data analysts from access to identifiable data

    h. At the German Institute for Cancer Epidemiology, data analysts are never given access to personal identifiers and can not access the computer system used by staff that process data.
    i. At the Swedish National Board of Health and Welfare, there is a specific statistical unit, the registry unit, which is permitted access to data containing identifying numbers. This unit cleans and processes the data and conducts data linkages for approved projects and de-identifies the data. Board analysts with permission to access files, see only de-identified data and never have access to the identified data

5. Restricting staff who process person identifiers from access to patient health records

    j. At the Singapore Ministry of Health, a very small number of employees in the ministry of health (3-4 people) have access to identifiable personal health data for linkage and these employees only see identifying variables. They never see the medical records associated with the patient identifiers.

6. Tracking and monitoring approved staff access to data

    k. At the Singapore Ministry of Health, staff analysing data must do so from within a secure data lab. The use of the data within the lab is monitored and if there was ever misuse of the data, it would be possible to identify the researchers involved.
    l. Employees of the Belgian Cancer Registry with access to identifiable data must have their access logged.
    m. At the Swedish National Board of Health and Welfare, a security officer tracks which employees have been granted access to data.

n. The National Board of Health in Denmark monitors who has access to registries and monitors and keeps logged how people with access are using the registry data on a 24/7 basis. The same protection and oversight applies to all national institutions in Denmark.
o. The U.K. NHS Information Centre regularly reviews access logs to ensure that employees are still using the files that they are approved to access. A similar monitoring has also been introduced at the Swiss Federal Statistical Office.

7. Training for new staff

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

8. Refresher training for existing staff

    r. The United States NCHS employees receive training on data security and confidentiality annually. Further, there are posters put up around the offices reminding staff about data confidentiality protection and security.
    s. The Belgian Cancer Registry provides training on global procedures regularly, including data security.
    t. The U.K. NHS Information Center requires employees to take online training each year in data protection and then to pass a test.
    u. Every two months, employees of the Republic of Korea's HIRA undergo data security and privacy training to ensure strict adherence to guidelines.
    v. The Canadian Institute for Health Information has a security month annually where employees attend in-person sessions.
    w. The U.K. NHS NSS Scotland has on-line training in data security that is scenario based.

9. Training for external researchers

    x. The Finland NIHW provides university-based researchers with a half-day training course on the NIHW databases, where part of the training is about data protection.
    y. The U.S. NCHS requires researchers with approved access to a Research Data Centre to take training on data security and confidentiality annually.

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

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

11. Secure buildings and offices

    aa. The German Institute for Cancer Epidemiology, where analysis of cancer registry data takes place at a national level, has strong physical security including doors that cannot be opened from the outside without a key. There is a clean desk requirement for staff engaged in data entry where no record can be left out at the end of the day. Records to be destroyed are stored in a separate container that cannot be easily accessed and a truck with a shredder comes monthly to security dispose of these materials.
    bb. The Swedish National Board of Health and Welfare stores data in a building that is locked and secure.

cc. At the Finland NIHW, individuals may only share an office with another staff member who has approved access to the same data.

12. Secure data transfer of identifiable data

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

13. Secure computer systems for the storage of identifiable data

hh. At the Singapore Ministry of Health, there is a two-card authentication system for a small number of authorized employees to gain access to identifiable personal health data. The computer system used to process the identifiable data and conduct data linkages is completely separated from the computer system for analysis of de-identified data. The identifiable data is further protected by a firewall.
ii. In Switzerland, IT security requirements are under a specific federal department (IT) and all federal data is centrally stored and protected. Physical displacement of data is avoided.
jj. In Sweden, identifiable databases of the National Board of Health and Welfare are not stored on computers that are connected to a network, which protects the data from unauthorized access.

14. Whole-of-government regulations or reporting up requirements on data security protection

kk. The U.S. has federal regulations on data security that federal agencies must follow. The U.S. NCHS must report to the government each year on its data security, and on any IT system changes that have occurred. The IT security is accredited every three years by the Centre for Disease Control. All federal agencies in the US would have a similar oversight and monitoring of their IT security.
ll. The Republic of Korea HIRA has internal guidelines on the protection of data security and confidentiality including specific guidelines related to data linkage. Under the requirements of the new Personal Information Protection Act, the National Information Service has issued guidelines on data security to government ministries including HIRA. HIRA will report annually to both the internal HIRA auditor and to the National Information Service on its data security.
mm. The Belgian Cancer Registry has privacy and information security policies and a data security plan required under the legislation authorizing the registry. This plan is updated every three years. Elements of the security plan include how and when access to data is permitted, including levels of access to personal health data.
nn. The NHS NSS Scotland data security respects British Standards for Information Management and Data Sharing and NHS Scotland standards.
oo. In Singapore, there are guidelines within government for data protection.

15. Third party or external data security audits

    pp. At the Belgian Cancer Registry, there are security audits by an independent organization that will attempt to attack the security of the registry. The registry has received a high rating by the independent organization for the results of its most recent security audit.
    qq. In the Republic of Korea, the Ministry of Public Information and Security and the National Information Service have the authority to conduct privacy and security audits of HIRA.
    rr. In Denmark, the Danish Data Protection Agency annually audits the National Board of Health to ensure that the handling of the databases meets legislative requirements. The Danish National Audit Office, which ensures that all national agencies comply with all relevant legislation may also audit the Board, or may rely on the results of Data Protection Agency audit.

16. Protocols in the event of a data security breach

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

17. Legal penalties for deliberate breaches of data security.

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

Data security when researchers receive data from public authorities

129. Data security is highest among data custodians requiring external researchers to access de-identified personal health data within a secure facility that is controlled by the data custodian. This practice was noted in the United States, Singapore and Canada (see section 4.2.3). As discussed earlier, many data custodians provide approved researchers with access to de-identified data. Below are several examples of how data security is approached in this situation.

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

131. In *Denmark*, the project approval will describe to the researchers the retention period of the file and will bind the researcher to not linking the data to any other databases and to not disclosing the data to a third party. The data protection authority in Denmark is then responsible for follow-up with the researchers to ensure compliance and data security audits take place. Non-compliance with the Act is a legal violation and subject to penalties. At the data destruction date, the researcher will be given the option to de-identify the data if they would like to retain the data for a longer period.

132. In the *United Kingdom*, the NSS Scotland indicates that the researcher is scrutinized during the approval process. A researcher who is a registered professional risks losing their profession as a result of a

deliberate breach and as a result would be more likely to be approved. A researcher working within a recognized institution where data protection and data security are known to be high would also be more likely to be approved. Researchers sign their application that binds them to data security; to data confidentiality protection (including following rules for vetting any tables intended for publication); and to not share the data they have received with a third party. The NHS Information Centre indicates that there have been cases where the National Information and Governance Board has requested that the linked data be given to a trusted third party for analysis, so that the risk of re-identification could be reduced.

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

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

Multi-country projects

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

136. The data protection legislations in some European countries make clear that it is possible to share identifiable data with other countries in the European Union. Noting this feature as part of national data protection legislations were both the *Denmark* National Board of Health and the *United Kingdom* NHS NSS Scotland.

137. The *United Kingdom* NHS NSS Scotland indicated that under the U.K. *Data Protection Act*, it is not acceptable to share de-identified individual data outside of the EU unless it can be demonstrated that the receiving country has the same standards for data protection as the U.K. Some non-EU countries have been certified as having equivalent standards and, for them, the process is the same as for an EU country. For a country not on the list, the two options for data access are a review of the country's legislation and an application for certification; or the provision of a fully de-identified data set, where there would be a very low risk of re-identification of individuals.

138. The *Denmark* National Board of Health has contributed de-identified individual data to multi-country studies with other Scandinavian countries and has provided aggregate study results to multi-country studies led by many other countries including France, the United Kingdom and Germany. Similarly, the *Finland* National Institute for Health and Welfare has participated in multi-country studies based on data linkages (see chapter 5).

139. The *Belgium* Cancer Registry may contribute de-identified individual-level data to a multi-country study if the Office of Data Protection grants permission. In *Singapore*, an international project requesting de-identified individual-level data may be approved. In the *Republic of Korea*, there is one example of a multi-country study conducted by a researcher in Singapore where HIRA contributed de-identified data.

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

141. There is a new EU-funded project, EuroREACH, where representatives from participating countries in Europe and outside of Europe with experience in conducing national data linkage studies are working together to develop a handbook. The handbook would support researchers within and outside of government in the launch of multi-country health services research based on data linkages. It will draw on best-practice country examples in establishing comprehensive systems of performance measurement in European countries, and in granting research access to patient-level data for the study of health services. It will also report on the person-level databases within countries that could support analysis and research and the steps required to produce population-based linked data sets and use them for multi-national health research projects (EuroREACH, 2011).

Multi-country project examples

This section is under development

142. EUropean Best Information through Regional Outcomes in Diabetes (EUBIROD) is a public health project funded by the European Union that aims to implement a sustainable European diabetes register to monitor diabetes complications and the health of diabetes patients (EUBIROD, 2011). EUBIROD is amalgamating aggregate data from 18 diabetes registries across Europe and it was challenging for the participants to find common ground where the local requirements for data security and privacy would be respected. The solution was the Best Information for Regional Outcomes or BIRO system (Di Iorio, C.T. et al., 2009). In BIRO, each disease registry provides aggregated data for their region with very little to no re-identification risk using an on-line data transfer system (See chapter 5). In working with participating countries, the conclusion of the EUBIROD team is that the sharing of de-identified data from diabetes registers would not be possible and still succeed in securing the participation of a large set of countries.

143. The Nelson trial is a randomised trial of the potential to use low-dose CT scans to screen at risk patients for lung cancer (van Klaveren et. al., 2009). Data are from Belgium and the Netherlands. The trial began in 2004 and is continuing until 2015. The world is waiting for the trial results because this is the only study where patients were recruited from population registries where it could be certain that those in the no-screening group indeed had not been screened. Results will have worldwide implications for health system policy regarding the uptake of and guidelines for lung-cancer screening. The decision on the data linkages necessary for Belgium's continuation of Nelson are pending approval by the Privacy Commissioner.

144. EuroHOPE, the European Health Care Outcomes, Performance and Efficiency project, is a new initiative funded by the European Union and coordinated by the National Institute for Health and Welfare in *Finland* to evaluate the performance of European health care systems in terms of outcomes, quality, use of resources and costs through data linkages. Participating countries all have the necessary health information infrastructure and legal framework to undertake the data linkages and include *Norway,*

Sweden, Scotland, regions in *Italy* and the *Netherlands*. For EUROHOPE, each participating country will link health care administrative databases for in-patient hospitalisations, pharmaceutical data, and cancer registry and mortality data in order to begin to generate indicators of the quality of hospital-based treatments across the whole cycle of care that would be comparable across the participating countries. The five focus areas for the development of these health care quality indicators are acute myocardial infarction, stroke, hip fracture, breast cancer and low birth-weight infants.

145. EuroHOPE aims to develop indicators that could be recommended to the EU for routine reporting, develop methods for international comparative health services research based on data linkages of person-level data; and inform about the policy-relevant drivers of health care quality, including treatment practices, use of medicines and new medical technologies, waiting times, financing, and the organisation of care. EUROHOPE is following the analytical model established by National Institute for Health and Welfare in *Finland* (see chapter 5).

Add remaining multi-country studies from the questionnaires.

CASE STUDIES OF LINKAGE ACTIVITIES

This chapter is under development and will be distributed as a room document at the HCQI Expert Group meeting on November 18, 2011.

146. This chapter presents case studies of recent data-linkage projects that were identified by the countries participating in this project as relevant to health and health care policy. Section 5.1 presents twelve case studies where the lead researchers for the studies participated in an in-depth follow-up interview. Study leaders provided insight into the project approval, protection of data privacy including patient consent and data security, access to data and data linkage, study results and future directions. Section 5.2 presents a summary of other projects identified by countries as important examples of the value to patient care and health policy of data linkages.

5.1 Case studies

5.1.1. Monitoring performance, effectiveness and costs of treatment episodes, Finland

5.1.2 Quality assessment of medical services, Republic of Korea

5.1.3 Quality and efficiency assessments of clinical guidelines, Sweden

5.1.4 Effectiveness and safety of breast cancer screening, Germany

5.1.5 Birth outcomes studies, United Kingdom

5.1.6 Pathways of care for stroke patients, Canada

5.1.7 Understanding life expectancy of a nation – the Swiss National Cohort, Switzerland

5.1.8 Understanding health care users and health outcomes through linkages, United States

5.1.9 Data linkage centre: Information Centre for Health and Social Care, United Kingdom

5.1.10Data linkage centre: Institute for Clinical and Evaluative Sciences, Canada

5.1.11Data linkage centre: Kaiser Permanente, United States

5.1.12Comparing diabetes outcomes across Europe– the EUBIROD project, Italy

PROGRESS, CHALLENGES AND OUTLOOK FOR THE FUTURE

147. In general, the outlook for the future is positive in terms of the opinions of the experts and researchers interviewed in this study toward their country's technical ability to undertake data linkages to monitor and report on the health of their people and the quality of their health care. There is nearly universal agreement that data infrastructures are growing stronger and more capable of supporting this type of work. Many were also of the view that some of the growing pains associated with working with data protection authorities to arrive at ways of working effectively together were passing and that the process for seeking approval and safely and appropriately undertaking data linkage studies was getting clearer on both sides. Nonetheless, many countries still face significant challenges.

Countries where it is becoming easier to use personal health data to monitor health and health-care quality

148. All participants to this study were able to express their opinion on whether it has become easier or harder to use personal health data to monitor health and health-care quality over the past five years. Respondents in Denmark, Malta, Singapore, and the United States felt it was becoming **easier**.

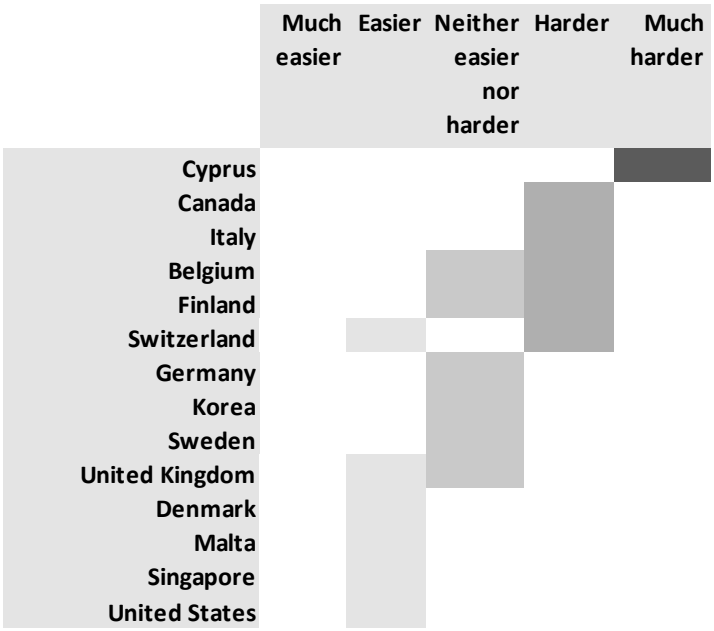
149. Strengths of the *Danish* information infrastructure for data linkage include its policy applications. Data linkage studies contribute to evidence-based decisions across a range of important policies from human resource planning for the number of doctors and medical specialists that will need to be educated today to meet the health-care needs of the population in the future; to how best to roll out population-based screening for cancer; to where to focus efforts to control rising health expenditures and the degree to which changes in tax rates could play a roll in improving the fiscal balance. Data linkages have helped to demonstrate the effectiveness of breast-cancer screening in real-world populations and to understand the effectiveness of drug and alcohol treatment approaches. Through data linkage, *Denmark* is able to evaluate options for consolidation of hospital services in terms of their impact on the population served.

150. Processes for working with external researchers have been simplified in *Denmark*. As early as four years ago, researchers were not aware that it was possible to request a linkage of their own data cohort. Now researchers are more aware that they may apply and how to do so. Further, researchers with on-going cohorts may apply for a running approval from the Data Protection Authority so that they may receive linkage results over a number of years without re-applying each year. An example would be a linkage to death certificates every year for several years for an on-going clinical cohort study.

151. The *United States* is behind other OECD countries in terms of infrastructure for health data linage. It does not have a unique patient identifier for health care encounters; there are so many different data custodians; there are multiple and complex laws regarding the use of personal health data; and the U.S. has been slow to implement E.H.R.s. One person was of the view that the U.S. is not meeting its responsibility for the public's health and that the population was unaware of the risks to their health that have resulted. For example, when an individual is in a health emergency, their care is similar to a battlefield response because their caregivers know nothing about them, including the medications they may be taking. Emergency response could be much safer with the secure sharing of medical records. There is a need to build awareness of the health consequences of not having a national health identifying number. For example, even among members of Kaiser Permanente, which has a high adhesion, people move in and out of the plan and their health records are incomplete. Medicaid recipients also move in and out of this plan and with a rolling back of eligibility for Medicare and Medicaid, there may be less information in the future. It is very difficult to understand health outcomes and health care quality as a result; and this

problem is the worst among the most vulnerable people because that is where long-term adhesion to particular insurance plans is the lowest.

Figure 3. Over the past five years, has it become easier or harder to use personal health data to monitor health and health-care quality?



Source: Sources: OCED HCQI Questionnaire, Secondary Use of Health Data, September and October 2011 and Follow-up Telephone Interviews, September and October, 2011

152. Two of the important barriers to a UPI in the *United States* are the public's trust in the use of their personal information in general, and the public's trust in government's use of personal information. Further, Congress, who allocates resources for federal agencies, is not always aware of the benefits to government of information from data linkages. This awareness is starting to rise with, for example, the opportunity to analyse survey data linked to Medicare and Medicaid records to identify the utilisation of care and the cost of care for vulnerable populations so that federal spending on these programmes can be allocated more efficiently.

153. In the *United States*, linkage is improving from a technical viewpoint. Computers are getting faster and there have been improvements in linkage methodologies that have made linkage projects easier to do. The analytical file from the linkage of population survey data to Medicare and Medicaid records at the NCHS is proving to be a challenging file to analyse. Methodological challenges result because the number of observations is affected by both the number eligible to link and the number eligible for government programmes and there is a project-by-project necessity to re-weight the data to correct for bias. An on-line tutorial is being developed to help researchers use the data.

154. As a result of the *Affordable Care Act*, there will be greater need to provide evidence of the impact of the *Act*. There are great expectations that data linkages will help to inform health policy, particularly in the area of effectiveness of care. If linkage studies in the next few years can show benefits to policy, data linkage will take off. There is a governmental push for E.H.R. systems. With many commercial software developers, there is a need to promote uniformity so that interoperability will be possible. On the horizon is also the use of genetic information from bio banking in long term epidemiological research. It may take twenty years or more of exposure to tobacco smoke, dust, and air

pollution to develop diseases. It therefore will take long-term linkage studies with genetic information, information on environment exposures, and information on health and health behaviours to know what factors are responsible for disease and to develop good policy responses. Future studies may involve a greater use of devices such as accelerometers to measure physical exertion or other devices to measure air quality and noise.

155. In *Singapore*, the data linkage projects that have taken place have had an impact on policy. At first, there were small sub-national studies and national efforts focussed on standards for data and data coding to improve the quality of person-level health databases. Once the quality of the databases was high, the door was then open for high quality data linkages to monitor the performance of the health system. For example, stroke care quality is not just about care at the time of the acute event, it is about seeing how patients are doing six months later. Similarly, linkages have permitted evaluation of the effectiveness of breast cancer screening. Through linkages there is better sight on blind spots at the national level that helps with a better assessment of how the health system is performing. For example, it is possible to monitor whether persons suffering from heart attacks were known to suffer from hypertension and diabetes and whether these conditions were being well managed in primary care.

156. It has taken time to develop a process for application and review of research proposals in *Singapore* that conforms to the legislative framework; and to arrive at the establishment of a secure data lab. It is on-going work to be respectful of privacy and to find the right balance between data protection and access. This is because new types of projects keep coming up that have never been considered before, including, for example, new projects where researchers may want cross-sectoral linkages. As new models of care develop, it will be important that there remains data available to follow a patient's care path. Singapore is developing a new consumer protection law. In the future it is possible that there may be creation of a central office for privacy issues. There is also a roll-out of a national E.H.R. and there is question of whether this could require additional legislation. There is potential to create confusion with too many pieces of legislation related to data protection. The introduction of E.H.R. introduces challenges to the coding of diagnoses, as the ICD coding system won't be sufficient. Singapore is considering SNOMED, but it is expensive. It will be important to preserve high quality data as the E.H.R. is implemented to ensure that coding is appropriate, so that the data quality is high and research results are valid. As more data becomes available in *Singapore* it may be necessary to consider a single body for conducting data linkages. The advantage would be economies of scale and standardization of linkage methods. Running data protection offices within each data custodian is expensive. A disadvantage could result from this approach, however, if it failed to service research needs in a timely fashion.

Countries where it is neither easier nor harder today than five years ago to use personal health data to monitor health and health-care quality

157. A second group felt that it was neither harder nor easier to use personal health data to monitor health and health-care quality today than five years ago (United Kingdom, Sweden, Republic of Korea and Germany). Views about the United Kingdom ranged from easier to neither harder nor easier.

158. The *Republic of Korea* notes that its strength comes from a unique identifying number that is used throughout the health-care system; a national system of health insurance that provides health care data for all patients; and a very strong technical infrastructure, where data is captured and stored electronically. The identifying number in Korea provides additional information to strengthen data linkages including full date of birth and place of birth. The conduct of data linkage has provided great benefits to Korea. Through analysis of health care claims, Korea is able to report on the quality of services provided by physicians, clinics, hospitals and long-term care. HIRA is also able to report on the cost of services and, with both quality and cost information, provide evidence for policy decisions. About ten years ago there was little discussion of protection of data privacy in the Republic of Korea. With the new legislation and increased

awareness, the balance between respect for patient privacy and the need for health research is sounder today.

159. It has never been easy to undertake data linkage studies in *Germany* and that situation remained unchanged over the past five years. While scientists have always been aware of the benefits of data linkage studies, there is a rising awareness among authorities of the benefits to policy. The mammography screening study that was mandated by law is a good example of this rising awareness (see chapter 5). Ten years ago, the introduction of a unique health information number in Germany would have been unthinkable. In October 2011, Germany announced that it would be moving forward with the introduction of a national HIN. In the future, it may be possible to undertake data linkages that are of a higher quality and at a lower cost; however, this will require approval in law. Another point of progress has been the introduction of German Cancer Aid, an organisation that awards scientific grants for record linkage studies. The existence of the awards is recognition of the scientific value of data linkages. Germany is doing well in the field of cancer registry in relation to some countries. Germany has cancer registry data for a population of 80 million people; federal legislation makes it possible to undertake data linkage studies with informed consent; and under some conditions without informed consent (although this is rare). The quality of data linkages based on pseudonomized names and other identifiers that is the basis of the German registry remains questioned by some. Certainly the probabilistic linkages are costly. It is a question of whether or not the changes that would be required to enable deterministic linkages in Germany would be supported by the population.

160. In *Sweden*, the coverage of the health care quality registers is better now than five years ago and it has not become harder or easier to undertake data linkage studies. The project to assess the impact of guidelines on processes of care and on patient health was requested by government and the results have been taken seriously. Because there are public quality reports for hospitals by name, the conduct of the assessment alone has lead to quicker adoption of care guidelines in hospitals (see chapter 5). There is increasing interest in the benefits of cross-sectoral studies in Sweden with data on social care and education to better understand the needs and the health outcomes of particular groups in the population and any differences in health care quality for different groups.

161. The Government of *Sweden* is considering new legislation regarding access to data and data linkages. The new legislation is to address the issue of commercial interests wanting to access personal health data. In particular, insurance companies are interested in using personal data to decide on when to approve or deny coverage. While this is an important issue, the concern is that the new legislation may have a negative impact on research that is in the public interest.

162. In the *United Kingdom,* there is greater interest in and political support for data linkage studies now than five years ago. There is recognition that these studies can meet needs for greater transparency about the quality of patient care and can improve health research and evaluation of outcomes of clinical trials. There is a new maternity strategy in Wales that has recognised there is not enough information on birth outcomes and infant health and pilot studies to evaluate whether data linkage could be used instead of primary data collection are leading to regular linkage programmes (see chapter 5). In the future, the linking of lab data and medical images will become possible and, in Scotland, a national database for the storing of radiology images is already in place.

163. Compared to five years ago, the establishment of the legislative framework and the creation of the National Information and Government Board (NIGB) as a governing body have helped to clarify for all researchers what is required to undertake a study and to provide a good mechanism to submit applications for consideration and approval. The new NHS Information Centre for Health and Social Care and the NWIS in Wales are now providing a services to facilitate data linkages and linkages will likely be used more in the future as they are much less expensive than primary data collection (see chapter 5). The Centre

does worry about the pressure on existing resources as the research community becomes more aware of its services.

164. From the perspective of researchers in the *United Kingdom*, it may seem that the approval process is long. It can take up to 6 months for a decision from NIGB and the Scotland NSS, indicates that the average time from submission of an application to a decision is three months with all finalized before six months. Resource constraints limit the Scotland NSS from being able to speed up the process. There have been a few instances of data loss in the U.K. that have raised public concerns about and interest in information governance that have made data security and confidentially rules tighter and processes for applicants wishing to access databases more difficult.

Countries where it is becoming harder to use personal health data to monitor health and health-care quality

165. A third group felt that it was becoming harder to use personal health data to monitor health and health-care quality (Canada, Cyprus[12], Italy, Belgium, Finland and Switzerland). Views were divided in Switzerland between easier and harder; while views about Belgium and Finland ranged from neither easier nor harder to harder.

166. In *Canada*, data sharing among public authorities is becoming increasingly complex as new legislations are introduced at both the provincial and federal levels that have an impact on the ability of the Canadian Institute for Health Information to receive data transfers from provinces to create national databases of individual-level data. The legislation in Quebec, which is preventing the transfer of identifiable personal-health information, restricts studies requiring data linkage to in-complete coverage of the nation. CIHI is working with Quebec to find a resolution. Within Ontario, the first province to introduce legislation specific to the protection of health data privacy, the legislation has helped to clarify consent requirements and end uncertainty that was limiting health research. The introduction of electronic health records will pose new challenges. From a legal viewpoint, the interoperability of electronic health records will be very challenging due to the privacy risks generated and the complexity of the Canadian laws.

167. The benefits of data linkage studies to public policy and patient care have been clearly demonstrated at the provincial level. The Institute for Clinical and Evaluative Sciences at the University of Toronto has published thousands of peer-reviewed scientific articles. Through data linkage at a population level, information is produced that informs about the effects of treatments in real-world populations with multiple morbidities which can differ from results of controlled trials. ICES results on the effectiveness of drug-eluding stents for heart patients surprised the medical community. There is evidence that research results have influenced policy and a good relationship with the Ontario Ministry of Health that appreciates that the study results help policy makers understand what can be done to improve the health care system. Further, there is rising interest among provincial policy makers in data to inform about the continuity of care and such information is made possible through data linkages.

168. There is growing interest in data linkages at the national level and a growing appreciation that data linkage adds information value to databases that is over and above their value as silos. The province of Ontario also reports a growing interest at the provincial level in comparing across provinces and that discussion is underway on how to move forward so that this type of pan-Canadian work could be done more effectively than is currently possible. There is also growing interest in cross-sectoral data linkages to, for example, understanding how health effects educational outcomes of young people or to understand when a province should place driving restrictions on elderly people with Alzheimer disease. The province

[12] See footnotes 1 and 2.

of Manitoba is leading the way for others in demonstrating the utility and importance of cross-sectoral linkages in Canada. Every time there is a media report of any type of data breach, however, there is a chill that sets back professional research organizations that are operating with strong data privacy protections.

169. *Canada* notes that there is an emphasis on knowledge translation to government from research work so that research results contribute to evidence-based decisions. The inverse knowledge translation, where governments help to clarify for researchers the legal requirements related to the use of personal health data needs the same attention. Within Canada, often people are saying the same thing but using different language and therefore not communicating clearly. There would be a benefit in developing clear definitions that are portable across provinces and in standardizing the interpretation of laws.

170. In *Cyprus*,[13] it has become harder over the past 5 years to conduct data linkages. Barriers include that there is no REB in Cyprus[14] that could hear applications for data linkage projects and to render a decision. Further, there is a general concern among data custodians that the sharing of identifiable data among governmental authorities is illegal. Under the European Directive, national governments can pass laws or regulations enabling the processing of identifiable personal health data without consent, but this has not happened yet in Cyprus[15]. The Ministry of Health may be the first, as it works toward new legislation to provide a legal framework to ensure that it will be able to continue to process death and cancer registry data. The European Commission has communicated that proposed changes to the data protection framework are forthcoming. What is needed is greater clarity in the provisions of the directive of when countries may process personal health data without consent.

171. Strengths of the *Italian* infrastructure include a large academic community and a long history of health and biological studies; established data flows for a spectrum of health services; universal health coverage which provides complete coverage of all patients in public data files; a unique identifying number that facilitates linkages; and an organization of care where each person is assigned to a physician which makes it much easier to study their care path. A number of new databases are being developed by the health ministry including cancer screening, emergency services and mental health services that offer the potential to improve population health monitoring at the national level.

172. One of the challenges for Italy is the fragmented nature of the administration of the health system. There is no adequate mechanism to share data across territories and provinces in Italy and sharing is nearly impossible, even for official institutions. Researchers seeking funding from granting agencies for projects where individual-level data would be needed from regions face great uncertainty about whether the project they have planned could be approved. This is because the criteria used by the regions to evaluate proposals are not known. The approval process is not transparent for those without a government partner and many researchers seek funding or collaboration with public authorities in order to have confidence that their project could be approved. For example, the National Outcomes Project is linking hospital and death records to develop more accurate indicators of deaths following treatment. The National Agency for Regional Services (AGENAS) is assisting and coverage is improving, but still the linkage is occurring in only a few regions. While some regions have technical problems, many are unsure if they can legally share de-identified data for a national project. Further the project is at risk from increasingly strict interpretations of privacy legislation that would only allow local authorities to link data for direct patient care. While regions have been authorized to conduct research and analysis with registry data, there is a growing concern that the Privacy Guarantor may revoke this approval. This concern has put a chill on health research in Italy as regions are becoming reluctant to participate in research studies. A concerning

[13] See footnotes 1 and 2.

[14] See footnotes 1 and 2.

[15] See footnotes 1 and 2.

development is the emergence of views in Italy that there should be an irreversible split between patient identifiers and the information about patient health and health care. Should these voices influence authorities, any data linkage of individual-level data would become impossible for both regional and national governments.

173. Overall it is more difficult to conduct linkages in *Italy* today than it was five years ago. A consequence is that policy decisions are lacking a strong evidence base. For example, the media report on cases of medical errors which alarm the public, but there is no national data on the extent of medical errors and whether the situation is improving or deteriorating. Policy focus is on expenditure control, and budget cuts may risk undermining health care quality or disease prevention. For example, the Abruzzo region, whose capital experienced a recent earthquake, has a large deficit and has experienced a sharp reduction in budget for health expenditures, including hospital closures and restrictions on pharmaceutical prescribing. This same region has published no report on the public health outcomes of this budget cut and lacks autonomous capacity to use its health information for public health monitoring.

174. *Italy* would benefit from clear guidelines from public authorities on the process to seek approval for a health research project and best practices in the processing of personal health information including data linkage. Greater transparency in procedures where information is shared with the public is needed, such as a check list available on a web-site. There is no office at the national level to fulfil this role. Further, if approval processes to link and analyse health data could be standardized among the regions, so that there was one process for approval in Italy, it would be a great improvement. Guidance from the EU and the OECD could make clearer organisational approaches to providing access to personal health data; and the advantages for and the rights of the population in conducting analysis based on linked data. This could better inform local, regional and national authorities in Italy.

175. *Belgium* reports doing well compared with some of the challenges faced by other cancer registries. The registry has been able to satisfy the requirements of the Privacy Commission, while at the same time, preserving the quality of the registry. The new E-Health Platform provides a helpful service at no cost to the registry. The time required to attend to all of the required procedures and to prepare applications for the Privacy Commission, however, creates an administrative burden that is costly in terms of human resources. The Privacy Commission takes about three months from the receipt of a submission to render a decision. Sometimes, however, questions are returned and another three months will be needed. Further, whenever an external researcher proposes a linkage involving the cancer registry, the Privacy Commission holds the registry responsible. The cancer registry must work with the researcher to prepare the application and then have the proposal vetted by its internal review board for scientific merit and must declare to the Privacy Commission that they would be willing to provide the data.

176. *Belgium* has received a huge benefit to public health of having a registry and being able to produce indicators of health care quality. Analysis of the registry has influenced policy decisions and published results have contributed to scientific research. A further benefit of linkages is that they avoid the need to ask too much of physicians providing data to the registry. Helping to reduce the burden on physicians is important; particularly as new disease registries emerge. Lastly, linkage and analysis of linked data provide new views of the quality of the data and reveal problems that otherwise would remain uncorrected.

177. *Finland* has invested in high-quality registers, has strong data protection legislation, and has a national identity number to facilitate linkages. Data linkages have had an impact on policy decisions in Finland. The PERFECT study on outcomes of hospital treatment in the year following the hospitalisation indicated that low birth-weight infants have higher mortality in non-university hospitals and a law was passed that all low birth-weight infants should be cared for in university hospitals. There have also been audits of lower-performing hospitals as a result of PERFECT study results (see chapter 5). In *Finland*,

there are plans underway to expand the current monitoring of the quality of hospital care to primary care, long-term care and social services.

178. Compared with five years ago, there is more bureaucracy around the preparation of record-linkage study proposals for approval by the Research Ethics Board (REB) and the time required to prepare the applications is not insignificant. The PERFECT project team is presenting a proposal to the REB almost every month, as any project that requires new data to be linked necessitates a new application to the REB.

179. *Finland* is challenged to keep its strength in data linkage studies. The legislation that enables the registries will need to be updated. The current legislation, from the late 1980s, enabled registries to grow and develop over time. For example, the legislation refers to data about health care activities, without narrowly specifying those activities. As a result, as new forms of care have emerged, such as outpatient care, the registries have evolved. The concern in Finland is that harmonizing with EU legislation may restrict the content of the legislation when it is revised. Other concerns are related to staff and resource cutbacks that may limit the NIHW. Thus far, the NIHW has been able to find ways to keep costs down for external researchers by entering into research collaborations at no cost, and only recovering the cost of staff time for very time consuming requests.

180. Compared with some European countries, *Switzerland* may be viewed as behind in terms of its infrastructure for data linkage. However, Switzerland is privileged to have a full population cohort study that does not exist in many other countries. In Switzerland, there is increased sensitivity of populations to the protection of privacy and this is reflected in more restrictive guidelines that have made data linkages more difficult today than five years ago. Further, a law is being developed to create a national cancer registry. This law is likely to clarify patient consent requirements and may set the course for linkages in Switzerland with stronger patient identifiers. Nonetheless, there is concern that long-standing studies, such as the Swiss National Cohort, could be negatively affected if a determination was made that any of the limited set of identifiers the cohort team uses now for probabilistic linkages are no longer legal (see chapter 5).

181. *Switzerland* is moving away from a questionnaire-based population census to an address registry. The data on the register will be updated annually and thus will provide much more up-to-date information on the population than census did. The address register will also have Social Security Numbers (SSN). SSN are available on health insurance and mortality data. If it were ever possible in Switzerland to use an encrypted SSN to conduct deterministic data linkages, there would be an important improvement in data quality and external researchers would be confident of linkages executed within government. The Switzerland FSO is considering amending the ordinance to its authorizing legislation to include collection of the SSN. This is as a result of a legal opinion of the Swiss Data Protection Office that the satisfaction of this condition would enable the FSO to use the SSN in data linkages. The Cancer Registry does not have SSN, however, and probabilistic linkages would continue to be necessary.

182. Data linkages create efficiencies and reduce the burden on health care providers. The FSO would like to extend current data linkage efforts from a focus on in-patient treatments to a focus on out-patient treatments in hospitals and day-surgery centres. This extension would increase the ability to monitor health care quality and would add valuable information that will likely increase interest in data linkage. Real statistical programmes are also more than just collections of data. The data needs to be made analysis ready with good information about the data and its quality; and the data elements included need to be of good quality and ready for use. The preparation of analysis-ready data is also part of the planning for the future of the health data programme.

CONCLUSIONS AND PRELIMINARY RECOMMENDATIONS FOR NEXT STEPS AT THE INTERNATIONAL LEVEL

183. This study indicates that national data infrastructures are improving across countries and the technical capacity to undertake data linkage studies is greater today than it was five years ago. There is optimism across countries for further improvements in infrastructure in the near future. This includes greater opportunity to conduct higher quality linkages due to improvements in the availability of unique patient identifying numbers; improvements in the availability of person-level data; improvements in the quality of data; and improvements in record linkage methods. Some experts have also expressed optimism about the potential to continue data linkage studies in the future due to having reaching a shared understanding with their data privacy officials of the requirements to respect principles of data privacy. This includes standardized processes for project approval, access to data and data security. These more optimistic experts tended to be from countries with more experience in, and demonstrated policy benefits from, data linkages for monitoring the quality of their health care system.

184. Other countries have weaker health information infrastructure at the national level. Many of these countries have decentralized the administration of health systems and have not reached a consensus within the country of how the levels of government could work together. Data from decentralized systems needs to be brought together to support national information infrastructure and capacity for data linkages at the level of the country. A principle challenge is the lack of clarity about the interpretation of legislations concerning the protection of data privacy at the national and sub-national levels. This includes the legality of data sharing among public authorities and the provision of de-identified data for research. The lack of clarity about the interpretation of existing legislation tends to push governments toward introducing yet more legislation which, at the worst, can lead to more confusion and uncertainty, and into unintended negative consequences for research that is in the public's interest.

185. Countries have provided evidence of the considerable effort they put in to protect data security and to safeguard personal health data from loss or deliberately malicious acts. Efforts were clearly demonstrated in this study related to project approval processes; internal data security; and de-identification of data and security measures for external researchers. Efforts to balance protection of data privacy and access to data for research are also clearly evident. New forms of whole-of-government approaches to project proposal review and data linkage services are very interesting developments. Not only do these help to standardize requirements and practices for both the government and external researchers, they have the potential to be more efficient. As many study participants noted, it is very expensive for data custodians to maintain an internal capacity for project vetting, undertaking data linkages, ensuring legal compliance and providing services to external researchers. These expenses are compounded in countries where there are multiple data custodians at the regional, state or municipal levels.

186. A particular worry across countries today is that legislative reforms that are on the horizon, or that may be stimulated due to the implementation of E.H.R. systems, may turn back the clock on the progress that has been made in enabling data linkages and providing access to linked data for research. A second worry is that changes in organization of care and the introduction of E.H.R. systems have the potential to set back the quality of the national databases, by creating holes in the health care pathway or lowering the quality of the data elements, such as the coding of diagnosis. Resource limitations, and not meeting expectations of timeliness, are worries among bodies that approve project proposals and among bodies that conduct data linkages on behalf of others.

187. There are opportunities to work together at an international level to help countries with weaker information infrastructure to grow stronger; and to support all countries in being able to maintain current capacity and to enhance information infrastructure in the coming years.

Preliminary recommendations

Continue mutual learning about the development of health information infrastructure to monitor the quality of health care

188. This study has shown that the development of health data infrastructure in countries is continually advancing. Would it be worthwhile to continue to monitor the development of health information infrastructure?

189. Monitoring could help to:

- Promote shared learning about advancements and challenges in the development and use of health data;
- Promote international comparability of data and data linkages; and
- Uncover new opportunities for the development of internationally comparable indicators of the quality of care.

7.1.2 Set international guidelines and standards

190. Given data protection legislations have grown from the same source, the eight principles of data privacy protection, there is likely more in common among legislations than there are differences. For members of the European Union, the European Commission issued a communication in 2010 that it will be reviewing the general EU legal framework on data protection to be translated into new legislation in 2011.

191. In addition to this legislative process, is it time for the OECD, ideally in close collaboration with the EU, to move forward from the eight principles of data privacy protection, to setting out what it means to respect those principles in terms of health information and research?

192. Questions where guidelines and standards would provide helpful answers include:

- Under what legal frameworks and under what conditions can public authorities share identifiable personal health data and de-identified data with one another?
- Under what legal frameworks and under what conditions can public authorities share identifiable and de-identified data…
  - With researchers in the non-government sector within country?
  - With researchers outside of the country?
- What are recommended international standards for de-identification of data and what might be the recommended data security requirements associated with different de-identification approaches?

- What are the key considerations a review board should deliberate upon when making a decision to approve a data linkage project?

- What would be a recommended template for a project proposal application?

- What are the best practices in communication with the public about processes for proposal submission, approval bodies and approval steps and data confidentiality and privacy protection?

- What are the best practices for data security and protection for data custodians?

7.1.3 Review existing legislations guiding health data privacy protection within countries, compare them, and make recommendations as to whether or not reforms are needed to meet anticipated future challenges to data privacy protection

193. New challenges in the near-term include:

- The implementation of E.H.R. systems,

- New ways to organize health care,

- New types of national data including lab results, images and bio-banks,

- The use ICTs in population health studies, and

- Cross-sectoral data linkages.

194. By working together, we could help countries to manage legislative reforms in a way that meets health data protection requirements while reducing the further proliferation of a spider web of legislations speaking to health data protection.

195. This study could become a background document for the joint workshop of the HCQI expert group and the OECD Working Party on Information Security and Privacy on health information privacy and public health and health services research in planning for May 2012. The workshop is an opportunity to engage data privacy authorities in a process to reach a clearer and more standardized interpretation of existing data privacy frameworks.

REFERENCES (NOT COMPLETED):

Academy of Medical Sciences (2011), A New Pathway for the Regulation and Governance of Health Research, United Kingdom, January.

Andrews, N. (2002), "The Value of Linked Data for Research into Surveillance and Adverse Events", Symposium on Health Data Linkage Proceedings, Australian Government Department of Health and Ageing, Sydney, 36-39.

Barrett, G., J.A. Cassell, J.L. Peacock, M.P. Coleman (2006), "National Survey of British Public's Views on Use of Identifiable Medical Data by the National Cancer Registry", BMJ 332(7549): 1068-1072.

Black, N. (2008), "Maximizing Research Opportunities of New NHS Information Systems", BMJ 336(7636): 106-107

Canadian Institutes of Health Research, (2002), Secondary Use of Personal Information in Health Research: Case Studies, Ottawa.

Challine, L.C., Pirard, L.L.P., and Grosclaude, P., (2003), "Epidemiological Surveillence of Cancer: Choosing a System for National Surveillence", Environnement, Risques et Sante, 2(2): 105-11.

Chappel, A. (2011), "Multi-payer Claims Database (MPCD) for Comparative Effectiveness Research", presented to the NCVHS Full Committee Meeting, June 16, 2011.

Churches, T. (2002) "The use of probabilistic record linkage, public key cryptography and

trusted third parties to improve the protection of personal privacy and confidentiality in disease registers and tissue banks". Symposium on Health Data Linkage Proceedings, Australian Government Department of Health and Ageing, Sydney, 57-61.

Commission for the Protection of Privacy, "Conference Conclusions", (forthcoming spring 2011), International Conference on Privacy and Scientific Research: from Obstruction to Construction, Brussels, November 22-23, 2010.

Conway, P.H., and J.M. VanLare (2010), "Improving Access to Health Care Data: The Open Government Strategy", Journal of the American Medical Association, 304(9): 1007-1008.

Davies, C. and Collins, R. (2006), "Balancing Potential Risks and Benefits of Using Confidential Data", BMJ; 33(3): 349-351

Di Iorio, C.T, F. Carinci et al. (2009), "Privacy Impact Assessment in the Design of Transnational Public Health Information System, the BIRO project", Journal of Medical Ethics, 35(12): 753-761.

Di lorio C.T., R. Andány et al. (2010), Privacy Impact Assessment Report, EUBIROD Consortium.

Di Iorio, C.T. (2010), "Cross-border Flow of Health Information: is 'Privacy by Design' Sufficient to Obtain Complete and Accurate Data for Public Health in Europe: The Case of BIRO/EUBIROD Diabetes Registers" presented at the Third European Public Health Conference, Amsterdam, November 12.

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

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

EUBIROD (2011), www.eubirod.eu/index.html, accessed October 25, 2011.

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

Frank, L. (2001), "Epidemiology: when an entire country is a cohort", Science, 287(5462). 2398-99.

Fraser, A. (2003) "Privacy and the Secondary use of Data in Health Research in Scotland", J Health Serv Res Policy, 8Suppl1:S1:12-6.

Gill, L.E. (1999), "Ox-Link: The Oxford Medical Record Linkage System", Record Linkage Techniques – 1997 Proceedings of an International Workshop and Exposition, National Academy Press, Washington.

Goldacre, M. (2002), "The Value of Linked Data for Policy Development, Strategic Planning, Clinical Practice and Public Health", Symposium on Health Data Linkage Proceedings, Australian Government Department of Health and Ageing, Sydney, 95-98.

Gunter, T.D. and N.P. Terry (2005), "The Emergence of National Electronic Health Record Architectures in the United States and Australia: Models, Costs and Questions", J Med Internet Res 7(1).

Habib, J.L. (2010), "EHRs, Meaningful Use and a model EHR", Drug Benefit Trends 22(4): 99-101

HM Treasury (2011), The Plan for Growth, London.

Holman, C.D., A.J. Bass, D.L. Rosman et al. (2008), "A Decade of Data Linkage in Western Australia: Strategic Design, Applications and Benefits of the WA Data Linkage System", Australian Health Review, 32(4): 766-777.

Jutte, D.P, N.L. Roos, M. Brownell (2011), "Administrative Record Linkage as a Tool for Public Health Research", Annu Rev Public Health 21;32:91-108.

Kalra, D., R. Gertz, P. Singleton, H.M. Inskip (2006), "Confidentiality of Personal Health Information Used for Research", BMJ 333(7560): 196-198.

Kelman, C.W., A.J. Bass, and C.D.J. Holman (2002), "Research Use of Linked Health Data – A Best Practice Protocol", *Aust NZ J Public Health*;**26**(3):251–255.

Linder, J.A., J. Ma, and D.W. Bates (2007), "Electronic Health Record Use and the Quality of Ambulatory Care in the United States", Arch Intern Med, 167(13): 1400-1405.

Mladovsky, P., E. Mossialos and M. McKee (2008), "Improving access to research data in Europe", BMJ 3(36) 287-8.

National Collaborative Research Infrastructure Strategy (2011), "Population Health Research Network", *www.ncris.innovation.gov.au/Capabilities/Pages/PopHealth.aspx*, accessed August 23, 2011.

National Institutes of Health (2011), "Institutional Review Boards and the HIPAA Privacy Rule", *www.privacyruleandresearch.nih.gov*, accessed October 22, 2011

OECD (2011), "Meeting of the Health Committee at the Ministerial Level: Revised Draft Communiqué", 20 September, DELSA/HEA(2010)10/REV4, section 7.

OECD (2010), "Improving Value in Health Care: Measuring Quality", OECD Health Policy Studies, OECD, Paris.

OECD (2009), OECD Policies for Information Security and Privacy, OECD, Paris.

Paulinski, D. (2011), "ASPE/CMS Multi-payer Claims Database", presented to Academy Health ARM, June 14, 2011.

Population Health Research Network (2011), *www.phrn.org.au*, accessed August 23, 2011.

Roos, L.L., V. Menec, R.J. Currie (2004), "Policy Analysis in an Information-Rich Environment", Soc Sci Med, 58(11):2231-41.

Safran, C., M. Bloomrosen, E. Hammond et al. (2007), "Toward a National Framework for the Secondary use of Health Data: An American Health Informatics White Paper", Journal of the American Health Informatics Association 4(1) 1-9.

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

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

Statistics Canada (2010), Privacy Impact Assessment for the Longitudinal Health and Administrative Data Initiative, Health Statistics Division, Ottawa.

Statistics Canada (2006), Occupational and Environmental Health Research Projects: A Descriptive Catalogue 1978 to 2005, Ottawa, 82-581XIE.

Tiik, M. (2010) "Rules and Access Rights of the Estonian Integrated e-Health System", Studies in Health Technology and Informatics, 156: 245-256.

United States General Accounting Office (2001), Record Linkage and Privacy: Issues in Creating New Federal Research and Statistical Information, 01(126SP).

United Nations Economic Commission for Europe (2009), Principles and Guidelines on Confidentiality Aspects of Data Integration Undertaken for Statistical and Related Research Purposes, UNECE, Geneva.

UKCRC R&D Advisory Group to Connecting for Health, (2007), Report of the Research Simulations www.ukcrc.org/publications/reports/ (accessed April 15, 2011).

Van Klaveren, R.J., M. Oudkerk et. al. (2009), "Management of Lung Nodules Detected by Volume CT Scanning", New England Journal of Medicine, 361:2221-2229.

Verschuuren, M., G. Badeyan, J. Carnicero et al. (2008), "The European Data Protection Legislation and Its Consequences for Public Health Monitoring: A Plea for Action", European Journal of Public Health, 18(6): 550-551.

Vetenskapsrådets, Rättsliga Förutsättningar för en Databasinfrastruktur för Forskning Vetenskapsrådet (The Public Access to Information and Secrecy Act – Aspects Relevant for the Construction of Databases), (2010) Vetenskapsrådets Rapportserie, Stockholm.

Walsh, D., M. Smalls and J. Boyd (2001), "Electronic Health Summaries – Building on the Foundation of the Scottish Record Linkage System", Stud Health Technol Inform, 84(Pt 2): 1212-1218.

Wiktorowicz, M.E., Lexchin, J., Moscou, K. et al (2010), Keeping an Eye on Prescription Drugs, Keeping Canadians Safe: Active Monitoring Systems for Drug Safety and Effectiveness in Canada and Internationally, Health Council of Canada, Toronto.

Willison, D.J., L. Schwartz, J. Abelson, C. Charles, et al. (2007), "Alternatives to Project-Specific Consent for Access to Personal Information for Health Research: What is the Opinion of the Canadian Public?", J Am Med Inform Assoc, 14(6): 706-12.

ANNEX A: SURVEY

Members of the OECD Health Care Quality Expert Group participated in a questionnaire to explore the potential, the barriers and the best practices to link existing data to inform about health and health-care quality. The questionnaire sought information about the general environment in countries for the secondary use of personal health data, as well as specific case studies involving the use of personal health data. The questionnaire also asked for the names of contact persons that could be invited to a follow-up telephone interview to learn more about the general environment for secondary use of personal health data and the specific case studies.

The questionnaire was developed by the OECD and was reviewed by 6 members of the HCQI Expert Group, 1 external expert, and 7 members of the OECD Secretariat. Table 1 provides a list of country representatives that responded to the questionnaire. Responses were received from August 31 through to October, 2011.

Table 1: Countries that responded to the 2011 HCQI Questionnaire on Secondary Use of Health Data

Country	Contact persons for the completion of the questionnaire
Belgium	Mr. Chr. Decoster, Director General FPS Health, Federal Public Service Health; Dr. L. Van Eycken, Director, Cancer Registry
Canada	Ms. Kira Leeb, Director, Health System Performance, Canadian Institute for Health Information
Cyprus[16]	Dr. Pavlos Pavlou, Coordinator, Health Monitoring Unit, Ministry of Health
Denmark	Dr. Niels Herman, National Board of Health
Finland	Dr. Päivi Hämäläinen, Director of Department, THL National Institute for Health and Welfare
Germany	Dr. Irene Keinhorst, Senior Advisor, Federal Ministry of Health; Dr. Christa Scheidt-Nave, Head of Division, Department of Epidemiology and Health Monitoring, Robert Koch Institute
Japan	Dr. Etsuji Okamoto, Senior Researcher, National Institute of Public Health
Malta	Dr. Sandro Distefano, Consultant in Public Health Medicine, Department of Health Information & Research, Health Division
Portugal	Dr. Paulo Boto, Consultant, Department of Quality in Health, Directorate General of Health
Republic of Korea	Ms. Sun Min Kim, Commissioner of Healthcare Quality, Health Insurance Review and Assessment Service
Singapore	Dr. Lim Eng Kok, Deputy Director, Healthcare Performance Group, Ministry of Health

[16] See footnotes 1 and 2.

Sweden	Dr. Max Köster, Senior Researcher, The National Board of Health and Welfare; Ms. Marie Lawrence, the National Board of Health and Welfare
Switzerland	Dr. Jacques Huguenin, Head of Health Care Statistics, Swiss Federal Statistical Office
United Kingdom	Ms. Alexandra Lazaro, Assistant Statistician, Department of Health; Ms. Anthea Springbett, Programme Principal, NHS NSS Information Services Division; Gavin Shivers, Health Statistics and Analysis Unit, Welsh Government
United States	Dr. Edward Sondik, Director, National Center for Health Statistics

OECD 2011 HCQI Questionnaire

Secondary Use of Health Data

Note: At the Health Care Quality Expert Group meeting of May 27, 2011, members agreed to participate in a questionnaire to explore the potential, the barriers and the best practices to link existing data to inform about health and health care quality. This questionnaire seeks information about the general environment in your country for secondary use of personal health data as well as specific case studies involving secondary use of personal health data. Results will be integrated into a report for the HCQI Expert Group meeting in November 2011 and will also contribute to a joint workshop of the Health Committee and the Working Party on Information Security and Privacy. For more background information see DESLA/HEA/HCQ/A(2011)5 and DELSA/HEA(2011)3.

For help with the questionnaire, contact Jillian Oderkirk by e-mail: jillian.oderkirk@oecd.org or by telephone: +33 1 45 24 76 03

We would appreciate your response by August 31, 2011. Please send completed questionnaires to jillian.oderkirk@oecd.org

Country:	

Please provide us with the contact information of the person primarily responsible for coordinating the completion of this questionnaire.

Name:	
Position:	
Organization:	
Address (postal):	
Email:	
Telephone:	

Please paste here using landscape the remaining pages of this questionnaire which is contained in the sheets Questionnaire part a, part b and part c in the file:
HCQI_Secondary_Uses_of_data_Questionnaire_September and October 12.xls

ANNEX B: TELEPHONE INTERVIEWS

The OECD conducted a series of telephone interviews with individuals identified within the country questionnaire responses as persons to contact to either discuss the country's current capacity to undertake health studies requiring the analysis and linkage of personal health data for public health and health services research or to discuss a specific project. In some cases, the same individual provided information on both the general environment and a specific project.

Table 1: Countries that responded to the 2011 HCQI Questionnaire on Secondary Use of Health Data

Country	Participants in a telephone interview
Belgium	Dr. L. Van Eycken, Director, Cancer Registry
Canada	Ms. Anne-Marie Phillips, Chief Privacy Officer, Canadian Institute for Health Information Ms. Cheryl Gula, Manager, Health Reports, Canadian Institute for Health Information Mr. Josh Fagbemi, Project Leader, Canadian Institute for Health Information Ms. Pamela Slaughter, Chief Privacy Officer, Institute for Clinical and Evaluative Sciences, University of Toronto
Cyprus[17]	Dr. Pavlos Pavlou, Coordinator, Health Monitoring Unit, Ministry of Health
Denmark	Ms. Anne-Marie Andersen, National Board of Health
Finland	Dr. Mika Gissler, Research Professor, THL National Institute for Health and Welfare Dr. Unto Häkkinen, Research Professor, THL National Institute for Health and Welfare
Germany	Dr. Alexander Katalinic, Director, Institute of Clinical Epidemiology/Institute of Cancer Epidemiology, University of Luebeck
Italy	Dr. Fabrizio Carincini, Consultant AGENAS and Technical Coordinator, EUBIROD Project, University of Perugia Ms. Concetta Tania Di Iorio, Serectrix snc
Japan	Ms. Natsuko Fujii, International Affairs Division, Ministry of Health, Labour and Welfare
Republic of Korea	Ms. Sun Min Kim, Commissioner of Healthcare Quality, Health Insurance Review and Assessment Service Mr. Yong Tai Ryu, Manager, Research Division, Health Insurance Review
Singapore	Dr. Lim Eng Kok, Deputy Director, Healthcare Performance Group, Ministry of Health and Dr. Lee

[17] See footnotes 1 and 2.

	Dr. Lee
Sweden	Dr. Björn Nilsson, Researcher, The National Board of Health and Welfare
Switzerland	Dr. Jacques Huguenin, Head of Health Care Statistics, Swiss Federal Statistical Office
	Dr. Adrian Spörri-Fahrni, Swiss National Cohort Manager, Bern University, Institute for Social and Preventative Medicine
United Kingdom	Ms. Xanthe Hannah, Section Head, NHS Information Center for Health and Social Care
	Dr. Janet Murray, Public Health Consultant and Caldicott Guardian, NHS Scotland
	Ms. Gwyneth Thomas, Statistician, Welsh Government
	Ms. Julie Messer, Principal Researcher, Health, Office for National Statistics, Wales
United States	Ms. Eve Powell-Griner, Confidentiality Officer, National Center for Health Statistics
	Ms. Jennifer Parker, Chief, Special Projects Branch, National Center for Health Statistics
	Ms. Donna Miller, Special Projects Branch, National Center for Health Statistics
	Dr. Mark Hornbrook, Chief Scientist, Kaiser Permanente

ANNEX C: TABLES

Table C1: Data available at a national level

	Hospital in-patient data	Primary care data	Cancer registry data	Prescription medicines data	Mortality data	Formal long-term care data	Patient experiences survey data	Mental hospital in-patient data	Population health survey data	Population census or registry data
Belgium	1	1	1	0	1	0	0	1	1	1
Canada	1	0	1	0	1	1	0	1	1	1
Cyprus*	1	0	1	0	1	0	0	0	1	1
Denmark	1	1	1	1	1	0	1	1	1	1
Finland	1	1	1	1	1	1	1	1	1	1
Germany	1	1	1	1	1	1	0	0	1	1
Japan	1	1	0	1	1	1	Nr	nr	1	1
Republic of Korea	1	1	1	1	1	1	1	1	1	1
Malta	1	1	1	0	1	1	0	1	1	1
Portugal	1	1	1	1	1	0	0	1	1	1
Singapore	1	1	1	0	1	1	0	0	1	1
Sweden	1	0	1	1	1	0	1	1	1	1
Switzerland	1	0	0	0	1	1	0	1	1	1
United Kingdom	1	1	1	1	1	1	1	1	1	1
United States	1	1	1	1	1	1	1	1	1	1
Total	15	11	13	9	15	10	6	11	15	15

Note: The data custodian should be a national authority and data should be included even when it does not cover 100% of the nation. 1-yes, 0-no, nr – no response
Note: *See footnotes 1 and 2.
Source: OCED HCQI Questionnaire, Secondary Use of Health Data, September and October 2011

Table C2: National data used to regularly report on health care quality

	Hospital in-patient data	Primary care data	Cancer registry data	Prescription medicines data	Mortality data	Formal long-term care data	Patient experiences survey data	Mental hospital in-patient data	Population health survey data	Population census or registry data
Belgium	1	1	1	na	1	na	na	1	1	Nr
Cyprus*	0	na	1	na	1	nr	na	nr	1	1
Canada	1	na	1	na	1	1	na	1	1	1
Denmark	1	1	1	0	0	na	1	1	0	0
Finland	1	nr	1	1	1	0	0	1	0	nr
Germany	1	1	1	1	1	1	na	na	1	1
Japan	nr	nr	nr	nr	nr	nr	nr	nr	nr	Nr
Republic of Korea	1	1	1	1	1	1	1	1	0	0
Malta	1	0	1	0	1	0	0	1	1	0
Portugal	1	1	0	1	0	na	na	0	1	Nr
Singapore	1	1	1	0	1	1	0	0	1	1

Sweden	1	na	1	1	1	na	1	1	1	0
Switzerland	1	na	Na	Na	1	1	na	1	0	1
United Kingdom	1	1	1	1	1	1	1	1	1	1
United States	1	1	1	1	1	1	1	1	1	1
Total	13	8	12	7	12	7	5	10	10	7

Note: The data custodian should be a national authority and data should be included even when it does not cover 100% of the nation. 1-yes, 0-no, nr – no response na – not applicable
Note: *See footnotes 1 and 2.
Source: OCED HCQI Questionnaire, Secondary Use of Health Data, September and October 2011

Table C3: National data containing records for patients (persons)

	Hospital in-patient data	Primary care data	Cancer registry data	Prescription medicines data	Mortality data	Formal long-term care data	Patient experiences survey data	Mental hospital in-patient data	Population health survey data	Population census or registry data
Belgium	0	1	1	na	1	na	na	0	0	nr
Cyprus*	1	na	1	na	1	na	na	na	1	1
Canada	1	na	1	na	1	1	na	1	1	1
Denmark	1	1	1	1	1	na	0	1	0	1
Finland	1	1	1	1	1	1	1	1	1	1
Germany	1	1	1	0	1	1	na	na	1	1
Japan	nr	nr	nr	nr	nr	nr	nr	nr	nr	nr
Republic of Korea	1	1	1	1	1	1	1	1	1	1
Portugal	1	1	0	1	0	na	na	1	1	nr
Malta	1	1	1	na	1	1	na	1	1	1
Singapore	1	1	1	na	1	1	na	na	1	1
Sweden	1	na	1	1	1	na	1	1	1	1
Switzerland	1	na	na	na	1	1	na	1	1	1
United Kingdom	1	1	1	1	1	1	1	1	1	1
United States	1	1	1	1	1	1	1	dk	1	1
Total	13	10	12	7	13	9	5	9	12	12

Note: The data custodian should be a national authority and data should be included even when it does not cover 100% of the nation. 1-yes, 0-no, nr – no response, na – not applicable, dk – don't know
Note: * See footnotes 1 and 2.
Source: OCED HCQI Questionnaire, Secondary Use of Health Data, September and October 2011

Table C4: National data contains a unique patient identifying number that could be used for record linkage

	Hospital in-patient data	Primary care data	Cancer registry data	Prescription medicines data	Mortality data	Formal long-term care data	Patient experiences survey data	Mental hospital in-patient data	Population health survey data	Population census or registry data
Belgium	1	1	1	na	0	na	na	0	1	nr
Cyprus*	0	na	1	na	1	na	na	na	1	1
Canada	1	na	1	na	1	1	na	1	1	1
Denmark	1	1	1	1	1	na	0	1	0	1
Finland	1	1	1	1	1	1	0	1	0	1
Germany	0	1	1	0	0	0	na	na	1	0
Japan	1	1	Na	1	1	1	nr	nr	1	1
Republic of Korea	1	1	1	1	1	1	0	1	1	1
Malta	1	1	1	0	1	1	0	1	0	1
Portugal	1	1	0	1	0	na	na	1	0	nr

Singapore	1	1	1	0	1	1	0	0	1	1
Sweden	1	na	1	1	1	na	0	1	1	1
Switzerland	1	na	Na	Na	0	1	na	1	0	0
United Kingdom	1	1	1	1	1	1	1	1	1	1
United States	0	0	0	0	0	0	0	0	0	0
Total	12	10	11	7	10	8	1	9	9	10

Note: The data custodian should be a national authority and data should be included even when it does not cover 100% of the nation. 1-yes, 0-no, nr – no response, na – not applicable, dk – don't know
Note: * See footnotes 1 and 2.
Source: OCED HCQI Questionnaire, Secondary Use of Health Data, September and October 2011

Table C5: National data contains identifying variables such as name, sex, birth date, and address that could be used for record linkage

	Hospital in-patient data	Primary care data	Cancer registry data	Prescription medicines data	Mortality data	Formal long-term care data	Patient experiences survey data	Mental hospital in-patient data	Population health survey data	Population census or registry data
Belgium	1	1	1	Na	1	na	na	0	0	nr
Cyprus*	1	na	1	na	1	nr	na	nr	1	1
Canada	1	na	1	na	1	1	na	1	1	1
Denmark	1	1	1	1	1	na	0	1	0	1
Finland	1	1	1	1	1	1	0	1	0	1
Germany	0	0	0	0	0	0	na	na	0	0
Japan	nr	nr	1	Nr	nr	nr	nr	nr	nr	nr
Republic of Korea	1	1	1	1	1	1	0	1	1	1
Malta	1	1	1	Na	1	1	na	1	0	1
Portugal	1	1	0	1	0	na	na	1	0	Nr
Singapore	1	1	1	Na	1	1	na	na	1	1
Sweden	1	na	1	1	1	na	0	1	1	1
Switzerland	1	na	Na	Na	1	1	na	1	0	1
United Kingdom	1	1	1	1	1	1	1	1	1	1
United States	1	1	1	1	1	1	1	1	1	1
Total	13	9	12	7	12	8	2	10	7	11

Note: The data custodian should be a national authority and data should be included even when it does not cover 100% of the nation. Identifying variables can include name, address, postal code, date of birth. 1-yes, 0-no, nr – no response, na – not applicable, dk – don't know
Note: * See footnotes 1 and 2.
Source: OCED HCQI Questionnaire, Secondary Use of Health Data, September and October 2011

Table C6: National data is used to undertake record linkage projects

	Hospital in-patient data	Primary care data	Cancer registry data	Prescription medicines data	Mortality data	Formal long-term care data	Patient experiences survey data	Mental hospital in-patient data	Population health survey data	Population census or registry data
Belgium	1	1	1	na	0	Na	na	0	1	nr
Cyprus*	0	na	1	na	1	Na	na	na	0	0
Canada	1	na	1	na	1	1	na	1	1	1
Denmark	1	1	1	1	1	Na	0	1	0	1
Finland	1	nr	1	1	1	1	0	1	0	1
Germany	0	0	0	0	0	0	na	na	0	0
Japan	1	1	na	1	nr	0	nr	nr	1	1
Republic of Korea	1	1	1	1	1	1	0	1	0	0

Malta	1	0	1	na	1	0	na	0	0	0
Portugal	0	1	0	1	0	na	na	0	0	Nr
Singapore	1	1	0	na	1	1	na	na	1	0
Sweden	1	na	1	1	1	Na	0	1	1	1
Switzerland	1	na	na	na	1	1	na	1	0	0
United Kingdom	1	1	1	1	1	1	0	1	1	1
United States	1	1	1	1	1	1	1	dk	1	1
Total	12	8	10	8	11	7	1	7	7	7

Note: The data custodian should be a national authority and data should be included even when it does not cover 100% of the nation. 1-yes, 0-no, nr – no response, na – not applicable, dk – don't know
Note: * See footnotes 1 and 2.
Source: OCED HCQI Questionnaire, Secondary Use of Health Data, September and October 2011

Table C7: National data is used to undertake record linkage projects on a regular basis

	Hospital in-patient data	Primary care data	Cancer registry data	Prescription medicines data	Mortality data	Formal long-term care data	Patient experiences survey data	Mental hospital in-patient data	Population health survey data	Population census or registry data
Belgium	1	0	1	na	0	na	na	0	0	nr
Cyprus*	0	na	1	na	1	na	na	na	0	0
Canada	1	na	1	na	1	nr	na	1	1	1
Denmark	1	1	1	1	1	na	0	1	0	1
Finland	1	na	1	1	1	1	0	1	0	1
Germany	0	0	0	0	0	0	na	na	0	0
Japan	0	0	0	0	0	0	0	0	0	0
Republic of Korea	1	1	1	1	1	1	0	1	0	0
Malta	1	0	1	na	1	0	na	0	0	0
Portugal	0	1	nr	1	nr	na	na	0	nr	nr
Singapore	1	1	0	na	1	1	na	0	1	0
Sweden	1	na	1	1	1	na	0	1	1	1
Switzerland	1	na	na	na	1	1	na	1	0	1
United Kingdom	1	1	1	1	1	0	0	1	1	1
United States	1	1	1	1	1	1	1	dk	1	1
Total	11	6	10	7	11	5	1	7	5	7

Note: A regular basis indicates that there is usually a project underway. 1-yes, 0-no, nr – no response, na – not applicable, dk – don't know
Note: *See footnotes 1 and 2.
Source: OCED HCQI Questionnaire, Secondary Use of Health Data, September and October 2011

Table C8: National record linkage projects are used for data quality monitoring

	Hospital in-patient data	Primary care data	Cancer registry data	Prescription medicines data	Mortality data	Formal long-term care data	Patient experiences survey data	Mental hospital in-patient data	Population health survey data	Population census or registry data
Belgium	1	0	1	na	0	na	na	0	0	nr
Cyprus*	0	na	1	na	0	nr	na	nr	0	0
Canada	1	na	nr	na	nr	nr	na	nr	nr	nr
Denmark	1	1	1	1	1	na	0	1	0	1
Finland	1	nr	1	1	1	1	0	1	0	1
Germany	0	0	0	0	0	0	na	na	0	0

Japan	0	0	na	0	nr	nr	nr	nr	nr	nr
Republic of Korea	1	1	1	1	1	1	0	1	0	0
Malta	1	0	1	na	1	0	na	0	0	0
Portugal	0	1	nr	1	nr	na	na	0	nr	nr
Singapore	1	1	1	na	1	1	na	na	1	0
Sweden	1	Na	1	1	1	na	0	1	1	Nr
Switzerland	0	Na	na	na	0	0	na	0	0	0
United Kingdom	1	0	1	0	1	0	0	0	0	0
United States	1	0	1	1	1	0	1	0	1	1
Total	10	4	10	6	8	3	1	4	3	3

Note: 1-yes, 0-no, nr – no response, na – not applicable, dk – don't know
Note: * See footnotes 1 and 2.
Source: OCED HCQI Questionnaire, Secondary Use of Health Data, September and October 2011

Table C9: Sub-national infrastructure for data linkage – regional or state-level record-linkage projects by type of data involved

	Hospital in-patient data	Primary care data	Cancer registry data	Prescription medicines data	Mortality data	Formal long-term care data	Patient experiences survey data	Mental hospital in-patient data	Population health survey data	Population census or registry data
Belgium	1	nr	nr	nr	0	nr	nr	0	dk	nr
Cyprus*	0	0	0	0	0	0	0	0	0	0
Canada	1	1	1	1	1	1	1	1	1	1
Denmark	na	Na	na	na	na	Na	na	na	na	na
Finland	na	Na	na	na	na	Na	na	na	na	na
Germany	0	0	1	0	1	0	0	0	1	0
Japan	nr	Nr	1	nr	0	Nr	nr	nr	nr	nr
Republic of Korea	na	Na	na	na	na	Na	na	na	na	na
Malta	na	Na	na	na	na	Na	na	na	na	Na
Portugal	0	0	1	0	0	Nr	Nr	0	0	Nr
Singapore	na	Na	na	na	na	Na	Na	na	na	Na
Sweden	1	1	1	1	1	1	1	1	1	1
Switzerland	nr	Nr	nr	nr	nr	Nr	Nr	nr	nr	Nr
United Kingdom	1	Nr	nr	1	nr	Nr	Nr	nr	nr	Nr
United States	dk	Dk	dk	dk	dk	Dk	Dk	dk	dk	Dk
Total	4	2	6	3	3	2	2	2	3	2

Note: 1-yes, 0-no, nr – no response, na – not applicable, dk – don't know
Note: * See footnotes 1 and 2.
Source: OCED HCQI Questionnaire, Secondary Use of Health Data, September and October 2011

Table C10: Sub-national infrastructure for data linkage – networks of health care organisations record linkage projects by type of data involved

	Hospital in-patient data	Primary care data	Cancer registry data	Prescription medicines data	Mortality data	Formal long-term care data	Patient experiences survey data	Mental hospital in-patient data	Population health survey data	Population census or registry data
Belgium	1	0	nr	nr	0	Nr	Nr	0	0	nr
Cyprus*	0	0	0	0	0	0	0	0	0	0
Canada	1	1	1	dk	1	1	dk	1	1	1

Denmark	**na**	**Na**	**na**	**na**	**na**	**Na**	**Na**	**na**	**na**	**na**
Finland	**na**	**Na**	**na**	**na**	**na**	**Na**	**Na**	**na**	**na**	**na**
Germany	**1**	**1**	**1**	**1**	**0**	**1**	**0**	**0**	**0**	**0**
Japan	**nr**	**Nr**	**nr**	**nr**	**nr**	**Nr**	**Nr**	**nr**	**nr**	**nr**
Republic of Korea	**nr**	**Nr**	**nr**	**nr**	**nr**	**Nr**	**Nr**	**nr**	**nr**	**nr**
Malta	**nr**	**Nr**	**nr**	**nr**	**nr**	**Nr**	**Nr**	**nr**	**nr**	**nr**
Portugal	**1**	**1**	**1**	**1**	**0**	**Nr**	**Nr**	**0**	**0**	**nr**
Singapore	**1**	**1**	**1**	**0**	**1**	**1**	**0**	**1**	**0**	**0**
Sweden	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Switzerland	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
United Kingdom	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
United States	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
Total	**5**	**4**	**4**	**2**	**3**	**4**	**1**	**3**	**2**	**2**

Note: 1-yes, 0-no, nr – no response, na – not applicable, dk – don't know
Note: * See footnotes 1 and 2.
Source: OCED HCQI Questionnaire, Secondary Use of Health Data, September and October 2011

ANNEX D: GLOSSARY OF TERMS

Term	Definition
Health data	Health data usually consists 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 [1] 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, behavioral patterns, sexual life, social and economic factors, etc. and healthcare administrative data: admissions and discharge data routine operational data, insurance and financial transactional data, etc.
Identifiable data	Data is identifiable if the information contains the name of an individual, or other identifying items such as birth date, address or geocoding. Data will be identifiable if the information contains a unique personal identifier and the holder of the information also has the master list linking the identifiers to individuals. Data may also be identifiable because of the number of different pieces of information known about a particular individual. It may also be possible to ascertain the identity of individuals from aggregated data where there are very few individuals in a particular category. Identifiability is dependent on the amount of information held and also on the skills and technology of the holder.
Database record	A database record is a row of data in a database table consisting of a single value from each column of data in the table. The data in the columns in a table 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.
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, Volume 1: Legal, Organizational and Technical Aspects, United Nations Studies in Methods, Glossary, Series F, No. 35, United Nations, New York 1991.)
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.

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.
De-identified information	This is information which does not identify an individual directly, and which cannot reasonably be used to determine identity. De-identification, also referred to as annonymisation, 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.
Confidentiality	Confidentiality relates to disclosure or nondisclosure of information. Historically a duty to honor 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 unauthorized disclosure.
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.
Data protection	Data protection refers to the set of privacy-motivated laws, policies and procedures that aim to minimize intrusion into respondents' privacy caused by the collection, storage and dissemination of personal data.
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.
Population census	A population census is the total process of collecting, compiling, evaluating, analyzing 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.
Network of health care organizations	A network of health care organizations provides a continuum of health care services. The network may provide integrated care under a parent holding company. Some networks have a Health Maintenance Organization (HMO) component.