Practices for the Review of Data Requests and the Disclosure of Health Information by Health Ministries and Large Data Custodians

Prepared for the Pan Canadian Health Information Privacy (HIP) Group

Authored by: Khaled El Emam, June 2010

Document Information

Document Title:	Practices for the Review of Data Requests and the Disclosure of Health Information by Health Ministries and Large Data Custodians
Document Date:	30 June 2010
Organization:	CHEO Research Institute, 401 Smyth Road, Ottawa, Ontario K1H 8L1, Canada
Contact:	Khaled El Emam (kelemam@ehealthinformation.ca)
More Information:	http://www.ehealthinformation.ca/
Sponsor:	Canada Health Infoway

Other Relevant Publications and Reports

K. El Emam: "Risk-based health data de-identification" In *IEEE Security and Privacy*, 8(3):64-67, 2010.

K. El Emam, A. Brown, P. AbdelMalik, A. Neisa, M. Walker, J. Bottomley, and T. Roffey: "A method for managing re-identification risk from small geographic areas in Canada" In *BMC Medical Informatics and Decision Making*, 10(18), 2010.

K. El Emam, F. Dankar, R. Issa, E. Jonker, D. Amyot, E. Cogo, JP. Corriveau, M. Walker, S. Chowdhury, R. Vaillancourt, T. Roffey, J. Bottomley: "A Globally Optimal k-Anonymity Method for the De-identification of Health Data ." In *Journal of the American Medical Informatics Association*, 16(5):670-682, 2009.

K. El Emam, A. Brown, and P. AbdelMalik: "Evaluating predictors of geographic area population size cutoffs to manage re-identification risk." In *Journal of the American Medical Informatics Association*, 16(2):256-266, 2009.

P. Kosseim and K. El Emam: "Privacy interests in prescription data. Part 1: Prescriber privacy." In *IEEE Security and Privacy*, January/February, 7(1):72-76, 2009.

K. El Emam and P. Kosseim: "Privacy interests in prescription data. Part 2: Patient privacy." In *IEEE Security and Privacy*, March/April, 7(2):75-78, 2009.

K. El Emam, F. Dankar, R. Vaillancourt, T. Roffey, and M. Lysyk: "Evaluating Patient Re-identification Risk from Hospital Prescription Records." In the *Canadian Journal of Hospital Pharmacy*, 62(4):307-319, 2009.

K. El Emam: "Heuristics for de-identifying health data." In *IEEE Security and Privacy*, July/August, 6(4):58-61, 2008.

K. El Emam, and F. Dankar: "Protecting privacy using k-anonymity." In the *Journal of the American Medical Informatics Association*, September/October, 15:627-637, 2008.

K. El Emam, E. Neri, and E. Jonker: "An evaluation of personal health information remnants in second hand personal computer disk drives." In *Journal of Medical Internet Research*, 9(3):e24, 2007.

K. El Emam, S. Jabbouri, S. Sams, Y. Drouet, M. Power: "Evaluating common de-identification heuristics for personal health information." In *Journal of Medical Internet Research*, 8(4):e28, November 2006.

K. El Emam: "Overview of Factors Affecting the Risk of Re-Identification in Canada", Access to Information and Privacy, Health Canada, May 2006.

Table of Contents

1 INTRODUCTION 4

2 METHODS 5

3 RESULTS 7

3.1 DATA HOLDINGS 8
3.2 DATA REQUESTS 11
3.3 REQUESTORS 15
3.4 HANDLING DATA REQUESTS 18
3.5 REVIEWS OF DATA REQUESTS 20
3.6 DISCLOSURE 22

4 DISCUSSION 26

4.1 KEY FINDINGS 26
4.2 LIMITATIONS 27

5 ACKNOWLEDGEMENTS 28

6 APPENDIX A: INVITATION EMAIL FOR PARTICIPANTS 29

7 APPENDIX B: INTERVIEW QUESTION GUIDE 30

8 REFERENCES 33

1 Introduction

It is often a challenge for government departments to exchange health information for secondary purposes. The challenge is exacerbated when these exchanges are inter-jurisdictional. This is partially driven by privacy concerns: in general, data custodians are reluctant to disclose personal health information for secondary purposes, especially when patient consent has not been sought for such secondary purposes.

Little is known about the practices used by health ministries in Canada for the review of data requests and for the disclosure of health information. Examples of unknown practices include how data requests are made by external parties, how they are reviewed, the criteria for deciding whether to disclose information, the de-identification practices used when the data is disclosed, whether these practices vary by the nature of the data requester/recipient and the data requested, and whether they are consistent within each ministry. Reasons why understanding such practices can be beneficial are that: (a) this may help us understand the data sharing barriers which exist, and (b) it may identify the practices that can facilitate more efficient inter-jurisdictional data exchange.

The purpose of this study was to produce a descriptive summary of the data review and disclosure practices of health ministries across the country. This study was undertaken to support the work of the Pan Canadian Health Information Privacy Group. We performed a series of interviews with key experts between April and June 2010, and the results of these interviews were summarized.

2 Methods

A basic model of the high level actors, activities, and data flows was constructed based on a review of the literature [1-4] and discussions with individuals responsible for the disclosure of health information for secondary purposes. This model was used to identify the open ended questions for the interviews, and is shown in Figure 1.

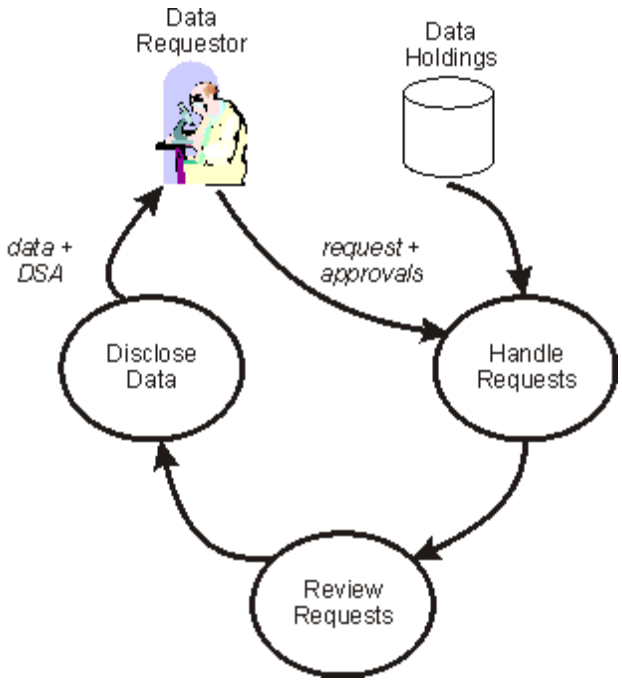

Figure 1: Workflow showing the various elements of the data request and disclosure processes that were being examined.

Interviewees at each ministry of health were nominated by provincial and territorial representatives. We also sought interviewees from large data custodians. These are organizations that were created to provide health information for secondary purposes. Sometimes these organizations have a narrow focus, such as providing data for research purposes only. In other cases these organizations will have a broader mandate which includes performing their own analyses for the provincial ministry of health. These large data custodians were included in our interviews because we expected them to have a different set of practices given their strong emphasis on disclosures for secondary purposes, and would provide us with a baseline to compare with.

The prospective interviewees were sent an invitation email (included in the appendix). If they accepted the invitation then that would be considered as consent to participate in the interview. Individuals who accept to be interviewed were contacted to arrange a convenient time to conduct a

telephone interview. During the scheduling process we also sent the interviewees a copy of the question guide to help them prepare.

The questions cover data requests for health information from other provincial and federal government departments, as well as academics, the media, clinicians, and the public. This will allow us to contrast how the practices differ by data requester.

The interview question guide was structured around four main practice categories: (1) data holdings, (2) data requests, (3) review, and (4) disclosure. It is included in the appendix. All interviews were digitally recorded and transcribed verbatim. The interviews ranged from 35 to 60 minutes.

The interviews were analysed using a combination of quantitative and qualitative content analysis [5, 6]. The initial coding of the transcripts was based on the categories in the general model shown in Figure 1.. However, throughout the analysis these practice categories were modified slightly to fit the data set. All transcripts were coded independently by three researchers at 80% agreement. Once all transcripts were coded, frequency counts were tabulated to determine how many participants in the sample referenced each specific practice category.

3 Results

A total of 22 individuals from 14 organizations were invited to participate in the interviews. Ten interviews were conducted with 13 individuals (59% response rate) from eight ministries of health and data custodians in both large and small jurisdictions (57% response rate) spanning the country from East to West.

The major practice categories presented in the interview data are summarized and discussed below. For each practice category, a frequency count is provided. It is important to note that this frequency count refers to the number of organizations who stated that the specific practice is currently in place and being implemented within their organization. We did not assess how well a particular practice was being implemented. If a practice was being performed regularly then it was counted. If a practice was planned or in pilot then it was not counted. Caveats and explanations to the counts are provided in the "Researcher Notes" column.

We also provide the *percentage* of organizations that have a practice implemented. The denominator of this percentage is the number of organizations for whom we had a response. This qualification is important because in some interviews it was not possible to ask about a particular practice, although this did not happen often. Therefore, the percentage is out of those who responded.

The "Researcher Notes" column provides more in-depth information and explanations on the following: (a) why the practice was deemed to be important, (b) qualifications and caveats for a count, (c) examples illustrating how a practice was being implemented, and (d) examples illustrating why a practice was considered not to be implemented or alternative practices that are in place that may mitigate a practice not being in place. We did not include quotes from the interviews because often these would reveal the identity of the respondent or would have to be truncated significantly that they lose important context.

While the coding we performed was subjective, two factors dilute the subjectivity: (a) three different coders extracted the information from transcripts, and this reduces the subjectivity in interpreting the transcripts, and (b) the "Researcher Notes" are intended to provide some of the context to interpret the counts.

It should also be noted that the results are presented anonymously to ensure that no specific custodian is identifiable. An assurance of individual and institutional anonymity was important to allow the interviewees to answer all of the questions in as much detail as possible.

3.1 Data Holdings

The purpose of this item was to understand how the organizations account for and handle their data holdings. While there was variation in the practices that were deployed, all of the organizations recognized that this was an important priority.

Practice	Number of Organizations (%)	Researcher Notes
Classify data holdings based on their sensitivity	1 (12.5)	• *A classification system for sensitivity is useful for allowing the organization to specify appropriate access controls and uses for each type of data. For example, highly sensitive information would have more stringent access controls in place.* • *Most organizations do not have a classification system. However, some responding organizations have made the point that because we are talking about health information, all the data is considered equally highly sensitive.* • *The exception was an organization which used the federal government's unrestricted/protected/confidential/restricted scheme. Although it seemed that the classification scheme was not widely applied in that case.*
Classify data holdings based on perceived identifiability	0 (0)	• *Another way to possibly classify data to be able to manage access controls is by their identifiability.* • *Organizations that had data at the individual patient level treated it all as identifiable data. There were controls on who has access to this identifiable data. However, this is not really a classification system per se because different levels of identifiability were not considered (i.e., there is only one classification level).*

Practice	Number of Organizations (%)	Researcher Notes
		• *Another distinction that is often made is between individual level data and aggregate data, whereby aggregate data is not considered personal information. However, this distinction pertains to the data that is disclosed as opposed to the data holdings themselves and therefore is not applicable here.*
Regularly maintain and update information about data holdings	7 (87.5)	• *The intention of this item was to understand whether there was a single centralized catalogue of all data holdings and whether this catalogue was being actively maintained.* • *There was considerable variation in how a catalogue of data holdings was maintained and how frequent this catalogue was updated. In some cases the interval between update cycles or planned update cycles was a few years. Some examples follow below.* • *Some organizations have an annual cycle to update their catalogue of holdings when that is a requirement, for example, of their board. Others have produced a catalogue at some point in the past but that catalogue has not been updated.* • *In some cases there is variation within the organization itself, with some service areas maintaining a catalogue themselves that may be paper or electronic. In such cases there is no central catalogue. Other organizations were in the process of creating such a catalogue as part of data warehousing initiatives.* • *Some organizations post their data holding information on their web sites. The amount of detail varies, from information about years covered, data owners, to a listing of*

Practice	Number of Organizations (%)	Researcher Notes
		all fields. • *In general, organizations who have a specific data custodian mandate (i.e., do not provide services to patients) and a fixed number of data sets are more able to produce such a catalogue and better able to keep it updated.*

3.2 Data Requests

This set of questions focused on the process by which data requests could be submitted to the data custodian, and whether that process varied by the type of data requestor and the type of data they were requesting.

Practice	Number of Organizations (%)	Researcher Notes
Receive initial requests for budgeting and scoping purposes	3 (37.5)	• *When someone wants to submit a data request for individual level data, some organizations will accept an initial short request to provide a scope and budget. This means the custodian will inform the requestor whether the data exists (i.e., whether the request is doable) and how much it would cost to get the data for them. Then the requester would get the funding and come back when they have the money to cover the costs. This is most often used in the context of requests from researchers who need a budget to include in their funding proposals. Then the researchers would come back with a full application once their funding has been approved.* • *In practice, many of the initially scoped & budgeted requests do not come back, possibly because it was not possible to get the funding or maybe because the question was answered through other means.*
Separate requests for aggregate versus individual level data	5 (62.5)	• *Some organizations separated the process for handling requests for aggregate versus individual level data. This means that there is a different process, and possibly even a separate group/person to send the request to.* • *Media requests are almost always for aggregate data.*

Practice	Number of Organizations (%)	Researcher Notes
		• *In other cases there was no real process for separating these two different types of requests or the custodian does not provide aggregate data.*

Practice	Number of Organizations (%)	Researcher Notes
Establish a clear and standardized definition of aggregate data to ensure that it is de-identified	0 (0)	• *The classification of a request as being for aggregate data is critical because these requests receive less scrutiny from a privacy perspective. Aggregate data is considered de-identified and are not regularly scrutinized by a privacy officer.* • *The definition of aggregate data was often not clear, and how it was assured that such data was truly de-identified was also not always clear. In most instances data that did not include names and addresses and that was in tabular form was considered aggregate data. In other cases experienced staff in the program area or in a decision support group were tasked with making that determination on a case by case basis.* • *In some cases data that was historically released as aggregate data continues to be (i.e., an analysis was done at some point in the past to determine that this was not identifiable data and therefore there is a precedent).*
Established guidelines for de-identification of data	2 (25)	• *Many custodians mentioned that record level data was typically de-identified prior to release, but in only two cases did custodians indicate that there are set guidelines for how this de-identification is performed.* • *The process is often reliant on an experienced person or group who understands the privacy risks and how to address them.* • *In some cases, the agency only possesses "de-identified" data stripped of direct identifiers , but would sometimes further generalize the data prior to disclosure based on the*

Practice	Number of Organizations (%)	Researcher Notes
		requirements of the requestor. • *Some custodians indicated that they would require patient consent for release of identifiable information.*

3.3 Requestors

This set of items was intended to understand who the typical data requestors were and the extent to which data requests from them were acceptable.

Practice	Number of Organizations (%)	Researcher Notes
Have pre-established, long-term data sharing arrangements with the requestor	4 (50)	• *Some data requests come under pre-established data sharing agreements, and these ones are straight forward to deal with. For instance, some provincial organizations have data sharing agreements with Health Canada or other federal agencies, and a data request may come in pursuant to that agreement. There are also requests from other government departments within the same province and there are pre-existing understandings about providing them the data for the purpose of providing a service to the population.* • *Where the data requests are internal within the same organization then they are dealt with in a different way than external requests. Because we are mostly focused on external requests here we did not collect information on processes for handling internal requests.* • *Some of the organizations that do not have any long-term data sharing arrangements with external parties are the ones established for handling data for research purposes only.*
Receive and accept data requests from individuals residing outside of the	6 (75)	• *In principle the data custodians would accept data requests from organizations outside their province (but within Canada). In practice, it is often not easy to do so, as*

Practice	Number of Organizations (%)	Researcher Notes
organization's province		*detailed below.* • *For some custodians without a formal process for making data requests, the ability to make a data request will depend on who you know at the data custodian. Therefore, this creates a natural barrier for out of province requesters.* • *Also, some institutions may charge considerable additional fees to out of province investigators.* • *Some custodians forbid any data, even de-identified data, from leaving the province. Researchers from out of province would have to travel to the province in question to perform their analysis.* • *In the case of researchers requesting data, some custodians require the researcher to come and work on site or to access the data through a secure connection only available within the province. Therefore, in such a case it is much easier to find a collaborator within the province who would have more ready access to the data.* • *Some custodians will accept out of province requests and will deal with them as they do other requests. For example, a request for research will require an REB approval, although sometimes the REB has to be local.* • *In practice, most data requests are from within the province.*
Receive and accept data requests from individuals residing outside of Canada	1 (12.5)	• *In principle custodians will process out of country data requests. However, this is relatively rare and it is not clear that such requests would be accepted.*

Practice	Number of Organizations (%)	Researcher Notes
		• *Generally, these requests would be subject to the same provisions as local requests.*
Accept data requests from industry-based organizations (e.g., pharmaceutical companies)	2 (25)	• *Some organizations indicated that it is possible for them to accept requests from industry on the condition that they meet the same standards as non-industry applications.* • *In the case of a request from a researcher, they may be funded by a commercial entity. However, in that case there are certain rules about the arrangement between the researcher and the commercial entity to ensure that there is no conflict, and then these are treated as regular requests.* • *Otherwise it is very rare to get requests for individual level of data from commercial entities, although they sometimes ask for aggregate data.*

3.4 Handling Data Requests

These items were intended to characterize the process by which data requests were handled by the custodian once they were received. It should be noted that where the process was not centralized, the interviewee was not always able to describe the process in other parts of the organization, but only the parts where s/he was involved or that were visible to him/her.

Practice	Number of Organizations (%)	Researcher Notes
Use standardized process or forms for requesting individual-level data	5 (62.5)	• *A standardized process means that there are specific forms that need to be completed and there are specific steps that an application will go through during the approval process.* • *Some custodians have the process explained clearly on their web site with the forms available on-line. This process tends to be better defined and developed when the data custodian has a narrow mandate, such as research, or when the custodian's main business is disclosing data for secondary purposes.* • *Some custodians do not have a centralized process. This means that a request may come in through a service or program area and it may get dealt with within that area. In such cases the request becomes known to a central privacy officer if questions are raised by the service or program area.*
Allow requestors to link identifiable data	0 (0)	• *None of the custodians provide sufficient identifiable information to allow the data requestor to link the data to some external database. If linking is required then special arrangements are done through trusted third parties, by the custodian, or not at all.*

Practice	Number of Organizations (%)	Researcher Notes
Perform data linkages for requestors	6 (75)	• *Some data custodians have multiple data holdings internally and will link them internally if linkage is approved.* • *In some instances the custodian will take a data set from the requestor and link it for them with their own data, and return the linked data.* • *In some cases linkage requires an additional level of approval beyond the regular data request process.*
Catalogue all received data requests	4 (50)	• *This item pertains to having readily available records for each request and how it has been handled and approved.* • *In general, there will be some records (e.g., of signed data sharing agreements), but these records do not always appear to be kept in a systematic fashion that would be easily accessible and queried.*

3.5 Reviews of Data Requests

These items were intended to understand the process by which data requests were reviewed by the custodian and who performed these reviews. The review criteria that were used were quite consistent across the interviewees, and included: whether the request was for data that was actually available, whether the data requested necessary to achieve the purpose, whether the appropriate approvals were obtained, and how the data can be de-identified and still be useful. The process of engaging with the data requester was iterative when the request was for individual level data. The number of iterations were largely a function of the complexity of the request. The duration of the review cycles varied considerably and there was no discernable pattern.

Practice	***Number of Organizations (%)***	**Researcher Notes**
Requests reviewed by privacy officer	*4 (50)*	• *Not all data custodians have a person designated as a privacy officer. Where one exists, that person would typically be part of the review process for all requests for individual level data (not necessarily aggregate data). In those cases a privacy officer would make the final call on whether data can be disclosed.* • *Some custodians have data access committees that review requests and these may substitute for or complement a privacy officer.*
Requests reviewed by data owner	*2 (25.0)*	• *Some data custodians are only holding and managing data on behalf of the data owner. In such cases an additional authorization from the data owner is required before the data can be disclosed to an external party.*
A formal review process exists	*6 (75.0)*	• *These custodians have put a process in place for handling data requests as they come in, with a defined point of entry and contact individuals.*

Practice	***Number of Organizations (%)***	**Researcher Notes**
		• *Some custodians do not have a formal process, in which case requests go to service or program areas directly or to individuals within the organization with direct access to the data who respond by themselves.*
Conduct a privacy impact assessment	*1 (16.7)*	• *Only one custodian made the explicit point that they can require a PIA to be performed if they deem that to be necessary.*
An appeals process exists	*0 (0)*	• *None of the custodians had a formal appeals process for data requests that they have denied. Although, denials are rare because the custodians would work with the data requestor to figure out what they need and negotiate an appropriate data set with them.* • *However, two custodians mentioned that a requestor can make a request under Freedom of Information to get access to the relevant documentation for their request and escalate the issue to an executive of the organization if they are not satisfied.*

3.6 Disclosure

These items describe what happens when a decision is made to disclose data to the requestor. This includes the immediate activities of making the data available as well as the longer term follow-up and monitoring of how the data was being used.

Practice	Number of Organizations (%)	Researcher Notes
Physical distribution of data	4 (50)	• *Some custodians provide physical data sets to their requestors either on encrypted CD or memory sticks (USB drives).*
Provision on-site access	2 (25)	• *Some custodians allow on-site access to their data. This may not be the exclusive means of access (i.e., they may provide remote access as well).*
Provision of remote access	2 (25)	• *Remote access allows data users to view and analyze data but not save it on their local machines nor print it. They would only have access to the data for which they have been approved. The remote access site may be the data requester's site or a designated remote site at specific locations.* • *Two additional custodians were piloting projects to establish remote access in order to make their data more easily accessible.*
Use of a secure file transfer system	4 (50)	• *Some custodians provide a secure file transfer system (sftp or equivalent) where they make the approved data available to the data requester and they can download it.* • *At least one additional custodian indicated that they were working toward establishing a secure file transfer system for*

Practice	Number of Organizations (%)	Researcher Notes
		their data.
Data sharing agreements for disclosing data	7 (87.5)	• *Data sharing agreements would normally only exist for the disclosure of individual level data rather than for aggregate data.* • *Many custodians have standard data sharing agreement for researchers. Data sharing agreements would not be needed where more general MOUs exist between organizations.* • *In some cases, the data sharing agreements may just be confidentiality agreements.*
Data destruction stipulations included in data sharing agreement	6 (100)	• *This issue was not addressed in every interview; however, in every interview in which it was addressed the custodian indicated that they did include some data destruction stipulations in their data sharing agreements.* • *Custodians typically stipulate a time limit for researchers to possess a data set, after which the data is to be returned or destroyed (proof of destruction to be provided).* • *Many custodians indicated that the period would vary dependent on the research project, and could range from one to seven years or potentially longer. At least one agency indicated that approval would be updated periodically for ongoing longitudinal studies.*
Conduct audits to check compliance	1 (12.5)	• *One custodian noted that they do audits to ensure that the data is used as it should be.* • *None of the other custodians perform compliance audits on the data recipients to ensure that they indeed have the*

Practice	Number of Organizations (%)	Researcher Notes
		practices in place that they had agreed to in the data sharing agreements. • *One of the main barriers to auditing was the cost of doing so.*
Recover costs for external requests	6 (75)	• *Some custodians charge fees for providing the data, although not all data requestors are charged (e.g., another provincial government department may not be charged but a researcher would be charged).* • *The effort to scope out a data request and provide a budget may not be recoverable because some of those proposed projects do not get funded.* • *Cost recovery often focuses on the cost of data retrieval but does not cover the cost of all of the analyses and privacy reviews that are performed. Therefore, the functions related to privacy are part of the overhead rather than the general cost recovery scheme.*
Breach notification provisions in data sharing agreement	4 (50)	• *Often where there is a data sharing agreement with the data recipient there is a clause requiring the recipient to notify the custodian if there is a breach, and not to attempt to contact any of the patients themselves. However, if the data is de-identified then it may not be possible for the data recipient to notify the patients in any case.*
Review of final outputs (e.g., reports, articles, presentations)	5 (62.5)	• *This item pertains to the analysis results being sent back to the custodian to review before publishing or broad dissemination.* • *The primary reason is to check for potential privacy*

Practice	Number of Organizations (%)	Researcher Notes
		breaches in the analyses and reports before they go public, but also to identify potentially newsworthy items (in a good and bad sense) that the custodian needs to know about in advance to prepare a response. • *Some custodians indicated that review was not always performed; it was dependent on the nature of the data used for the analysis (e.g.. individual-level data vs. aggregate data).*

4 Discussion

4.1 Key Findings

Based on our analysis of the interviews, the following general conclusions can be drawn:

- In most cases the basic elements of a good data review and disclosure process were in place. In situations where a critical practice was missing this was recognized and there was an effort to put it in place. The existence of health privacy legislation in the jurisdiction, or the imminent introduction of one, is helping put these practices in place.
- Inter-jurisdictional data disclosures are difficult. Even where a custodian is open to such disclosures, practicalities make them difficult to operationalize (for example, remote access is only allowed within province).
- Data custodians with an explicit mandate to manage health data and make data available for secondary purposes, and that do not have service or program responsibilities tend to have more of the practices in place, and defined and repeatable processes for handling data requests. For example, custodians created specifically for making data available for research purposes would fall into that category.
- It seems that aggregate data is often interpreted to mean tabular data. The assumption that such aggregate data is de-identified needs further scrutiny. Tabular data, even with cell sizes less than five suppressed, may have quite a high risk of re-identification. Additional effort to have explicit criteria for determining when data is aggregate would be beneficial.
- All data custodians perform de-identification on the individual level data before it is disclosed. However, this sometimes results in full dates of birth and relatively accurate geographic information to be disclosed, which can have a high risk of re-identification. Additional effort to have explicit criteria for the de-identification of individual-level data would be beneficial.
- Performing audits on data recipients is quite rare. Data recipient enthusiasm about compliance to the privacy and security practices tends to decline over

time from the point of receiving the data. Audits, or a real threat of them, can help ensure that strong compliance remains over time.

- Efforts to centralize data, for example, through data warehousing programs, would make it easier to catalogue data holdings, to have a single process for handling data requests, and for disclosing data.
- Little attempt is made to recover the total cost of processing data requests. Unless disclosures for secondary purposes are part of an organization's mandate, investments in the activities described in this report have to come from other sources.

4.2 Limitations

Not all of the provinces and territories were covered during our interviews. Within the time available it was sometimes difficult to schedule a time with the appropriate person in some provinces. Consequently, it is plausible that important elements were missed. Although geographically we were able to cover most of the country.

It was not always possible to collect documents from the interviewees, either because they were not readily available or they could not be shared. This made it difficult to gather additional supporting details for some of the information collected during the interviews.

5 Acknowledgements

We wish to thank the provincial representatives who worked with us during this project to identify potential interviewees, and the interviewees for their time. Also, Katherine Moreau and her team for transcribing and coding the transcripts, and Elizabeth Jonker who coded the transcripts as well as coordinating the interviews.

6 Appendix A: Invitation Email for Participants

INVITATION EMAIL

Subject: Interview Study on the Disclosure of Health Information

We are conducting a research study to understand the practices used by health ministries and large health data custodians in Canada for the review of data requests and for the disclosure of health information. This will consist of a series of interviews of key health ministry staff. The results of this study will help us identify areas where new tools, templates, and methods may need to be developed that can potentially facilitate more effective and efficient data review and disclosure practices.

Your name was provided to us as someone who would have good knowledge of the data disclosure practices of your organization. We wanted to invite you to participate in this interview study on review and disclosure practices. The interview will last for one hour and will be conducted either by telephone or face-to-face at a time that is convenient for you. The interview will be recorded and later transcribed for analysis. Your identity and any information you provide will remain anonymous and confidential, and no raw interview data will be shared outside of our research team. All data will kept in a secure electronic vault.

This study was sponsored by Canada Health Infoway. This research protocol has been approved by the Children's Hospital of Eastern Ontario (CHEO) Research Institute Research Ethics Board (REB). You may contact the Chair of the REB for additional information (613) 737-7600 x3272.

If you are interested in participating in this interview study, please click on the following link:

http:/www.ehealthinformaiton.ca/R+D/interview?XXXXXX

You will receive a confirmation email and a research coordinator from our team will contact you to schedule a convenient time to conduct the interview.

Thank you very much for your consideration.
Khaled El Emam
University of Ottawa / CHEO Research Institute
(www.ehealthinformation.ca)

7 Appendix B: Interview Question Guide

OPEN ENDED INTERVIEW QUESTIONS

The following questions pertain to the key elements of the data request, review, and disclosure process. We are only concerned with discretionary disclosures of health information by your organization, rather than mandatory reporting requirements. We are also not concerned about disclosures that are part of public health reporting or of access or freedom of information requests.

In all responses to these questions, we would greatly appreciate examples to illustrate the responses. In addition, if there is any documentation to support the response that can be shared with us, it would be greatly appreciated.

Data Holdings

- Do you have a list or information about the key personal health information holdings within your ministry/registry (i.e., is there a central catalogue) ? If so, who maintains that, is it regularly updated, and how do you ensure that it is complete ?
- How are data holdings classified in terms of their sensitivity ? Are there centralized classification criteria, and who performs this classification ?
- Do you have authority to disclose that data – how would do you characterize this authority (e.g., from legislation, patients provided express consent to disclose the data for specific purposes) ?
- Do you have access to or manage data holdings at agencies or other organizations that you sponsor or fund ?

Data Requests

- How and how often do you receive requests for data from bodies external to your organization (e.g., other government departments within and outside your jurisdiction, academic organizations, clinicians, the media, commercial entities, and the public) ?
- How and how often do you receive requests from outside your jurisdiction ?

- How and how often do you receive requests for data from within the organization ?

- How and how often do you receive requests for data related to research, analysis, monitoring and evaluation ?

- Are all requests for data catalogued somewhere or does each department/group deal with request separately and maintain their own catalogues and systems ?

- How are requests for data made (e.g., do you have an Internet-based system for data requests, or do people send in paper requests, or by email) ?

- Do you receive requests to link data within the organization or link with data outside the organization ?

Review

- Who reviews requests for data within your organization ?

- What is the process used to decide whether a data request will be fulfilled ? How is that dependent on the type of data requested ? How is that dependent on whether the request is from within or outside your jurisdiction ?

- What factors are considered important in reviewing a request (e.g., cost, sensitivity of the data) ?

- Is that process dependent on the requester of the data (type and jurisdiction) ?

- Do you anonymize data before disclosure ? How is that done ? Are there any documented procedures ?

- How long does it take, on average, to review requests for data and make a decision on disclosure ?

- How are reviews documented ?

- Is there an appeals process for reviews that the data requester does not agree with ?

Disclosure

- For all of the questions below, how do the responses vary for recipients within and outside your jurisdiction ?

- How do you provide data to the different types of recipients (e.g., do you have a portal, on CD, is the data encrypted, sent via FedEx tapes to the

requester's site) ? Do you provide remote access to data for some data requesters ? Do you allow some data requesters to come on site to access the data ?

- Do you have standard data sharing agreements for disclosing data (or multiple standard ones for different types of data and recipient) ?
- Do you stipulate procedures for handling data breaches at the data recipient site ?
- If the data recipient is a public body, do you stipulate procedures for handling access/FOI requests which pertain to your data?
- Do you stipulate procedures for the data recipient restricting/expanding the purposes of the original disclosure ?
- Do you stipulate procedures/restrictions on the data recipient subsequently disclosing the data to another party for the same or a different purpose ?
- Do you perform audits on the data recipient to ensure that they comply with your stipulations for disclosing the data ?
- Do you require certain privacy/security practices at the data recipient site ?
- Have you every had to prematurely terminate a data sharing agreement – if so, why ?
- How often are disclosures sequential/incremental (i.e., data is provided to the recipient on a regular basis) ?
- How do you recover the costs for the data disclosures ?
- Do you require that the data recipient report to you or inform you if they make any portion of the data public or produce a public report from that data ?

8 References

1. Black C, McGrail K, Fooks C, Baranek P, Maslove L. *Data, Data, Everywhere -- Improving access to population health and health services research data in Canada*. 2005; Centre for Health Services and Policy Research and Canadian Policy Research Networks.

2. Paulette Collins, Pamela Slaughter, Noralou Roos, Karen Weisbaum, Marie Hirtle, Jack Williams, Patricia Martens, Laupacis A. *Privacy Best Practices for Secondary Data Use; Harmonizing Research & Privac*. 2006; Institute for Clinical Evaluative Science.

3. Mackie C, Bradburn N. *Improving access to and confidentiality of research data: Report of a workshop*. 2000; Washington: The National Academies Press.

4. Jabine T. *Statistical disclosure limitation practices of United States statistical agencies.* Journal of Official Statistics, 1993; 9(2):127-454.

5. Morgan DL. *Qualitative content analysis.* Qualitative Health Research, 1993; 3:112-121.

6. Weber RP. *Basic content analysis*. 1990: Sage.