1-1-2013

Registry data storage and curation

Megan Johnston
University of Calgary

Craig Campbell
Western University, craig.campbell@lhsc.on.ca

Rachel Hayward
Office of the Information and Privacy Commissioner of Alberta

Mark Lowerison
University of Calgary

Vanessa Noonan
Rick Hansen Institute

See next page for additional authors

Citation of this paper:
Johnston, Megan; Campbell, Craig; Hayward, Rachel; Lowerison, Mark; Noonan, Vanessa; Pfister, Ted; Maxwell, Colleen; Fortin, Claire; and Willison, Eric, "Registry data storage and curation" (2013). *Paediatrics Publications*. 2446.
https://ir.lib.uwo.ca/paedpub/2446

Authors

Megan Johnston, Craig Campbell, Rachel Hayward, Mark Lowerison, Vanessa Noonan, Ted Pfister, Colleen Maxwell, Claire Fortin, and Eric Willison

Registry Data Storage and Curation

Megan Johnston[1,7], Craig Campbell[2], Rachel Hayward[3], Mark Lowerison[1,7], Vanessa K. Noonan[4], Ted Pfister[1,7], Colleen Maxwell[5], Claire Marie Fortin[6], Eric E. Smith[1,7], Jean K. Mah[1,7], Moira K. Kapral[8], Nathalie Jette[1,7,9], Tamara Pringsheim[1,7], Lawrence Korngut[1,7]

Can J Neurol Sci. 2013; 40: Suppl. 2 - S35-S40

The storage of patient and medical information in a disease registry is a critical concept for consideration during registry design and development. The choice of data storage methods may influence the ability to access data in the future; the ability to store data long-term; and the ability to exchange data with other registries or research projects as required. Additionally choosing a data storage method involves a certain degree of uncertainty in an era that has gone from the file cabinet to the five inch floppy to the cloud in a matter of 35 years. In preparing this section of the guideline we reviewed available scholarly and grey literature resources; consulted with disease, registry, legal, ethics, privacy, and information technology (IT) experts; and consulted appropriate legislation and policy documentation in Canada.

Relevant Literature

Unfortunately our efforts to examine relevant literature in this topic area were unsuccessful. While there is a large body of IT literature on topics that may apply here, very little is specific to the Canadian context or the disease registry context. Where general principles applied we have reflected this as much as possible. Additionally, in some registry literature where mention to the issues of Data Storage and Curation were made we have noted this.

Policy and Legislation

Many Canadian provinces and territories have specific legislation components that address information technology applications and criteria that must be met by applications collecting health information. As a result, disease registry projects need to consider their relevant legislation within the jurisdiction in which the database itself will be housed, and any other additional needs that could be demanded of the registry based on the other jurisdictions in which it operates. Table 5 features a list of relevant documentation by province.

When examining software products to determine the best fit for a registry application; evaluate the product specifications to ensure that all legislative requirements can be met. Table 6 outlines some of the common requirements for neurological registries in Canada and some software products available in 2012 that meet some or all of the requirements.

Other Considerations

Storage Considerations

The type of database selected for a disease registry project will depend on a number of factors determined early in the registry development including: the expected number of records (database size); the expected number of users (database clients); the expected duration of the registry (length of data storage); the type of data being stored (data type); and the duration of the data storage after the registry project is complete. For example, in Canada, clinical trial data are required to be stored for 25 years under Part C Division 5 of the Food and Drug Regulations [C.05.012], however little consideration is typically given to the format of the storage of clinical trial data and whether or not this will remain accessible 25 years in the future. With electronic data storage, such considerations must not be underestimated. If registries are capturing both observational and clinical trial data, there may also be a need to store the observational data much longer than might normally be the case or to have the registry modules separated so that data from clinical trials can be stored for the longer time frame. These considerations should be made in advance of registry set up as they may impact the type of consent provided by patients in the area of data storage.

In addition to the above considerations disease registry projects may also want to consider Canadian legislation and privacy considerations with respect to data storage location. The following aspects should be considered:

A) Server Model (e.g. single server, dual server, or cloud server/storage?)
B) Physical Location of Servers (e.g. country, province, institution)
C) Physical Server Access (e.g. controlled, secure?)
D) Network location of Server (e.g. secured, visible, access controls)

From the [1]University of Calgary, Calgary, Alberta; [2]Western University – London Health Sciences Centre, London, Ontario; [3]Alberta Office of the Information and Privacy Commissioner, Edmonton, Alberta; [4]Rick Hansen Institute, Vancouver, British Columbia; [5]University of Waterloo, Waterloo, Ontario; [6]Canadian Institutes of Health Information, Toronto, Ontario; [7]Hotchkiss Brain Institute, University of Calgary, Calgary, Alberta; [8]University of Toronto, Toronto, Ontario; [9]Institute for Public Health, University of Calgary, Calgary, Alberta.

Final Revisions Submitted January 28, 2013.

Correspondence to: Lawrence Korngut, Clinical Neurosciences, 480060, 4th Floor Administration, South Health Campus, 4448 Front Street SE, Calgary, Alberta, T3M 1M4, Canada. Email: Lawrence.korngut@albertahealthservices.ca.

https://doi.org/10.1017/S0317167100017157 Published online by Cambridge University Press

E) Database user access (e.g. data access permission levels; authentication mechanisms)
F) Hardware and Software security controls (e.g. firewalls, encryption)

Database Genres

When selecting a database genre (database type) consider the complexity of the data processing that will be required during registry operation and the organizational resources available for the management of the database. Table 7, adapted from Brian Westrich, University of Minnesota[151] may be a useful tool during these considerations.

Following identification of the required database genre it will be necessary to select a specific software product with which to execute the database. Considerations during this process will include organizational assets (e.g. institutional licenses or IT services); budget (consider using open source software products if budget is small) and the development timeline. Additionally considerations must be made regarding the software product's ability to meet data storage requirements associated with legislation (See Table 5). Finally a key consideration during this stage involves the database size. The larger and more complex the database, the more important it becomes to select a software product that can create an efficient and readily accessible database while optimizing storage space. To this end, one must consider the structure of the database created by each database genre product. Additionally storage space will be impacted by the format of the data stored in the database.

Table 5: Relevant Legislation and Policy Relating to Software Considerations

Province/ Territory	Best Practice/Guidelines Document
Alberta	Personal Information Protection Act (PIPA) PIPA Advisory #8 Implementing Reasonable Safeguards (http://www.oipc.ab.ca/Content_Files/Files/Publications/PIPA_Advisory_8_Reasonable_Safeguards2007.pdf) [111]
	Alberta Electronic Health Record Regulation (http://www.qp.alberta.ca/documents/Regs/2010_118.pdf) [112]
	FOIP Guidelines and Practices Chapter 8. Records and Information Management (http://www.servicealberta.ca/foip/documents/chapter8.pdf) [113]
	Developing Records Retention and Disposition Schedules (http://www.rimp.gov.ab.ca/publications/pdf/SchedulingGuide.pdf) ([114]
	Health Information Act Guidelines and Practices Manual (http://www.health.alberta.ca/documents/hia-guidelines-practices-manual.pdf) [115]
	FOIP Guidelines and Practices (http://www.servicealberta.ca/foip/resources/guidelines-and-practices.cfm) [116]
British Columbia	Physicians & Security of Personal Information (http://www.oipc.bc.ca/tools-guidance/guidance-documents.aspx) [117]
	Information Management and Information Technology Management (http://www.fin.gov.bc.ca/ocg/fmb/manuals/CPM/12_Info_Mgmt_and_Info_Tech.htm) [118]
	FOIPP Act Policy and Procedures Manual (http://www.cio.gov.bc.ca/cio/priv_leg/manual/sec30_39/sec30.page?) [119]
	Information Security Policy (http://www.cio.gov.bc.ca/local/cio/informationsecurity/policy/isp.pdf) [120]
Manitoba	University of Manitoba Safe Computing Topics (http://www.oit.umn.edu/safe-computing/topics/index.htm) [121]
	Respecting Privacy: A Compliance Review Tool for Manitoba's Information Privacy Laws: A Special Report http://www.ombudsman.mb.ca/pdf/Special%20Report%20English%20CRT%20-%20Oct%207.pdf [122]
New Brunswick	Guidelines for Custodians to assess compliance with the Personal Health Information Privacy and Access Act (PHIPPA) (http://www.gnb.ca/0051/acts/pdf/7133%20%E2%82%AC%20English%20long%20list%203s.pdf) [123]
Newfoundland	The Personal Health Information Act (Resources) (http://www.health.gov.nl.ca/health/PHIA/) [124]
Nova Scotia	Personal Health Information Legislation for Nova Scotia (http://novascotia.ca/dhw/phia/custodians.asp) [125]
	Privacy Impact Assessment Template (http://www.gov.ns.ca/just/IAP/_docs/Appendix%20B%20PIA%20Template.pdf) [126]
Nunavut	
Ontario	IPC Ontario Privacy and Confidentiality When Working Outside the Office (http://www.ipc.on.ca/images/Resources/up-num_20.pdf) [127]
	Manual for the Review and Approval of Prescribed Persons and Prescribed Entities (http://www.ipc.on.ca/images/Findings/process.pdf) [128]
Prince Edward Island	According to the Forms and Resource Materials section of the Information and Privacy Commissioner's website (http://www.assembly.pe.ca/index.php3?number=1013951),[129] PEI's FOIP Act is based on Alberta's FOIP Act and cites the Guidelines and Practices: 2009 Edition (http://www.servicealberta.ca/foip/resources/guidelines-and-practices.cfm)[116] as a useful reference. Chapter 9 lists technical safeguards.
Quebec	Minimum Requirements for the Security of Computerized Records of Health and Social Services Network Clients (http://www.cai.gouv.qc.ca/documents/CAI_G_securite_doss_info_rsss_eng.pdf) [130]
	Exigences minimales relatives à la sécurité des dossiers informatisés des usagers du réseau de la Santé et des Services sociaux http://www.cai.gouv.qc.ca/documents/CAI_G_securite_doss_info_rsss.pdf [131]
	CAI Quebec (http://www.cai.gouv.qc.ca/english/) (http://www.cai.gouv.qc.ca/)
Saskatchewan	Saskatchewan Archives Board Records Management Policies and Guidelines (http://www.saskarchives.com/services-government/record-management-policy-and-guidelines) [132]
	Security Controls for Protection of Personal Information (http://www.justice.gov.sk.ca/ITOSecurityControlsforProtectionofPersonalInformation.pdf) [133]
	Government of Saskatchewan Resources and Tools Security (http://www.justice.gov.sk.ca/AP-Security) [134]
Northwest Territories	GNWT INFORMATION TECHNOLOGY Electronic Information Security (http://www.fin.gov.nt.ca/documents/ocio/ppsg/6003.00.27%20-%20Standards%20-%20Electronic%20Information%20Security.pdf) [135]
Yukon	ATIPP Compliance Assessment (http://www.ombudsman.yk.ca/uploads/general/ACA_ATIPP_Compliance_Assessment_August_2011.pdf) [136]
	Yukon Information and Privacy Commissioner Privacy Breach Checklist (http://www.ombudsman.yk.ca/uploads/general/ATIPP_Privacy_Breach_Checklist_2011.pdf) [137]
	Yukon Information and Privacy Commissioner Best Practice: Responding to a Privacy Breach (http://www.ombudsman.yk.ca/uploads/general/ATIPP_Best_Practice_Privacy_Breach_Response.pdf) [138]
	Privacy Impact Assessment (http://www.ombudsman.yk.ca/uploads/general/PRIVACY%20IMPACT%20ASSESSMENT.pdf) [139]
Canada	Operational Security Standard: Management of Information Technology Security (MITS) (http://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=12328&section=text) [140]
	Guidance Document: Taking Privacy into Account Before Making Contracting Decisions (http://www.tbs-sct.gc.ca/atip-aiprp/tpa-pcp/tpa-pcp06-eng.asp)[141]
	Electronic Health Record (EHR)Privacy and Security Requirements Reviewed with Jurisdictions and Providers (https://knowledge.infoway-inforoute.ca/EHRSRA/doc/EHR-Privacy-Security-Requirements.pdf) [142]
	Health Canada Final Audit Report – Audit of Information Technology (IT) Security (http://www.hc-sc.gc.ca/ahc-asc/pubs/_audit-verif/2011-04/index-eng.php#_Toc2008) [143]
Other	ISO/IEC 27002:2005 Information technology — Security techniques — Code of practice for information security management (http://www.iso27001security.com/html/27002.html#) [144]

https://doi.org/10.1017/S0317167100017157 Published online by Cambridge University Press

Table 6: Software features by product

FEATURE	Open-Source	Local Server Install	Authentication/ Platform	Password Controlled User Level-Authentication	User Action Log	Data encryption	Interacts with Third-Party Data Sources	Workflow Management	Patient Portal
PRODUCT									
REDCap (Produced by Vanderbilt University) www.project-redcap.org [146]	✓	✓ (if you are a participating site)	LDAP, Shibboleth or Table-Based (user selected)	✓	✓	Can be installed using additional software products	✓		
ClinicalPursuit www.patientregistrysoftware.com [147]		optional	Microsoft.NET	✓	✓		✓		
i2b2 – Informatics for Integrating Biology and the Bedside www.i2b2.org [148]	✓	✓	AJAX	✓	✓		✓	✓	
Patient Crossroads www.patientcrossroads.com [149]				✓					✓
Axiom Clarinet www.certus-tech.com [150]		✓	J2EE	✓				✓	

Database Structures

The type of database structure that is selected will influence many factors impacting the operation of the registry. These factors might include: computer hardware infrastructure; registry stability and performance; data entry and recall speed; and reporting capability.

Relational databases – This type of database (see Figure 2) is still a very common format created by many software products on the market however it can come with some significant limitations if data sizes are large.[152] These databases store data in a defined record where the common location of the data elements contained within the record is the sole logic between the data elements within the record. This limits the granularity of the database to the record level (i.e. data cannot be examined within a record except if the full record is recalled). As a result processing time to read and write records is high; total disk storage required for the database is high; and modifications to records require the whole record to be rewritten.

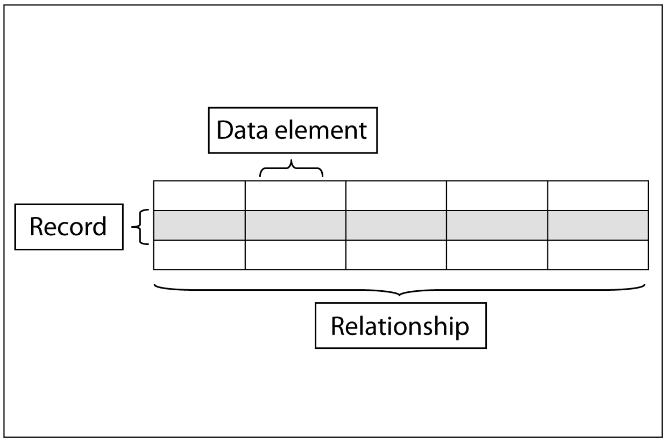


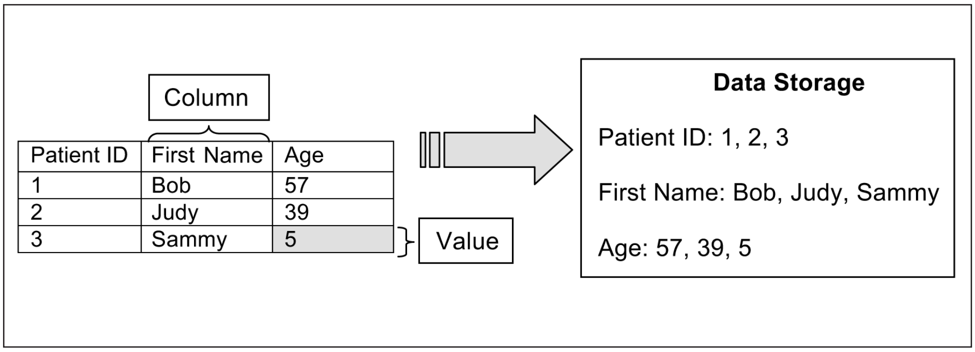
Figure 2: *Relational Database Structure*

Table 7: Database Genre Decision Support Tool

Complexity	Minimum required data resources	Database genre
Storage only	Paper case report forms	Non-automated file storage (paper based and/or electronic copies of paper forms).
Electronic storage	Computer	Word processor or basic file storage software. Basic file backup to external media (CD-ROM or DVD).
Structuring – Data that is stored needs to have different “fields” or “pieces”.	Semi-skilled staff	Spreadsheet. Note that this is still a storage only task and no analysis is required.
Relating – Data is stored in fields and there is a need to define relationships or examine relationships between the fields.	Computer staff (part-time)	Personal database tools (e.g. Access). These tools feature simple data form and query design tools. Multiple data tables can be created and relationships between them can be defined. Analysis required is simplistic.
Complex, high volume – There will be large amounts of data between which complex relationships exist. This may also involve the need to have simultaneous access by multiple users.	Computer staff (full-time)	Industrial database tools (e.g. Oracle, mySQL). These tools allow for all of the features of Personal database tools but also allow for logging of user transactions; simultaneous access and updates by multiple users and complex query construction (for example, construction of data sub-sets). These software products may also allow database architecture to span over multiple servers for operation and storage.
Highly specific or specialized – The type of data being collected; the data collection process and/or the queries and analysis required of the data require customization beyond that available in standard tools.	Highly skilled computer programmers. High performance computing equipment. This type of solution may also require custom networking.	Programming languages (e.g. Java, C-plus). These tools may operate in conjunction with industrial database tools or other library structures to fully enable the required database architecture.

https://doi.org/10.1017/S0317167100017157 Published online by Cambridge University Press

Columnar databases – This type of database (see Figure 3 below) is increasingly adopted due to the increase in analytical simplicity found through this method when compared to relational databases. These databases store information by column with all values within a column being stored as a single dataset (i.e. these datasets are made up of data from multiple "records").[152] A key advantage of this format is that "parts of records" from a relational database perspective can be analyzed and written or rewritten. This feature increases the speed with which data processing can be accomplished. However, the trade off here is that recalling records requires the assembly of data values across multiple columns into a pre-determined format which if the number of columns is large (complex dataset) or the number of requests is large (many users) may impact database performance.



Patient ID	First Name	Age
1	Bob	57
2	Judy	39
3	Sammy	5

Data Storage

Patient ID: 1, 2, 3

First Name: Bob, Judy, Sammy

Age: 57, 39, 5

Figure 3: *Columnar Database Structure*

Correlation databases – This type of database (see Figure 4 below) stores each data value only once and then stores references allowing collocation of appropriate values for each "record" using descriptive metadata. Like a card catalogue, metadata stores information on what values are required for each "record" and where each value can be found allowing programming that reads the metadata to reassemble each "record" when required. These databases have similar advantages to columnar databases in terms of partial record access and writing actions. However due to the low storage volume of correlation databases, their performance often exceeds columnar databases.[152]

Once the database genre and database structure are selected, the final considerations are the database formatting and the configuration of adequate backup infrastructure.

Data Formats

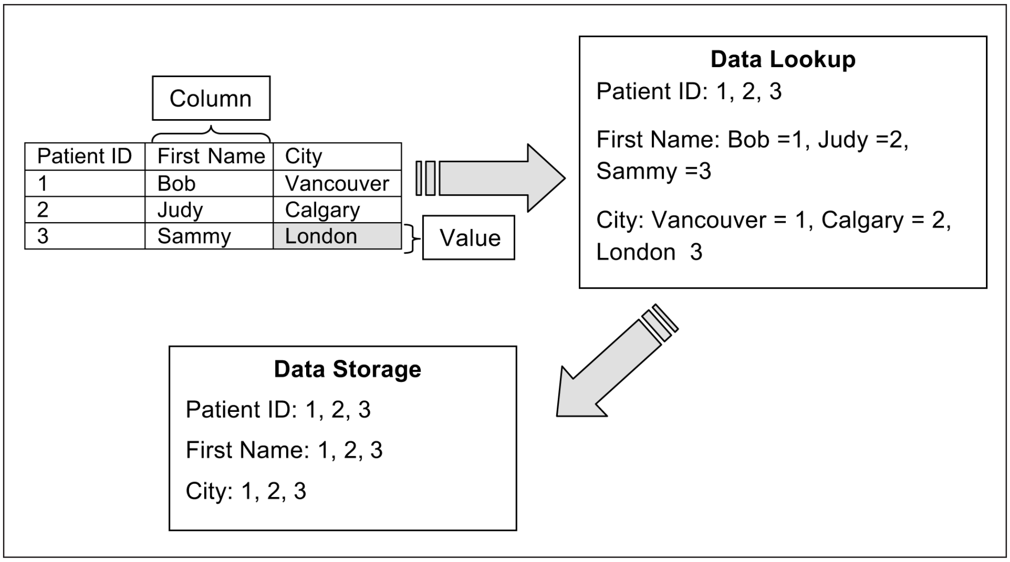
There are three ways to physically store digital data available on the current market[153]:

1) Magnetic storage (e.g. magnetic tape or hard drive)
2) Solid state (e.g. flash memory)
3) Optical (e.g. Blu-Ray, DVD, CD-ROM)

Table 8 discusses some of the considerations for each method. Clearly demonstrated by the information in Table 8, the choice of backup infrastructure is best addressed by choosing multiple physical storage formats. It should be a regular practice to perform backups at pre-determined intervals to a hard drive space and then periodic backups to an optical format.

In addition to physical storage format, in registry development one must consider file format obsolescence. This is the state during which the digital format of the file is no longer readable due to changes in technology and file formatting practices. File format obsolescence is independent of physical storage format and is to do with the actual digital format of the files on the physical storage format. Both are important considerations when storing data long term.

As it is impossible to define the file formats of the future and indeed to define file formats that will not go obsolete this guideline instead recommends a risk assessment approach to addressing this concern. This risk assessment approach is derived from File Format Obsolescence Risk Decision Support System (Version 1.1 released November 2007).[155]



Patient ID	First Name	City
1	Bob	Vancouver
2	Judy	Calgary
3	Sammy	London

Data Lookup

Patient ID: 1, 2, 3

First Name: Bob =1, Judy =2, Sammy =3

City: Vancouver = 1, Calgary = 2, London 3

Data Storage

Patient ID: 1, 2, 3

First Name: 1, 2, 3

City: 1, 2, 3

Figure 4: *Correlation Database Structure*

https://doi.org/10.1017/S0317167100017157 Published online by Cambridge University Press

Table 8: Data Format Considerations

Data Format	Life Expectancy	Pros	Cons
Magnetic tape	0.7 – 1083 years depending on storage temperature, humidity, availability of error-correction coding[153]	• Readily available • Cheap • Convenient • Sizeable	• Oldest technology • Many known impacts on life expectancy • Reliability depends on manufacturing
Hard disk drives	Limited knowledge available but may range from as early as 3 months independent of utilization[154]	• Readily available • Convenient • Sizeable	• Limited capacity (currently in terabyte (TB) range) • Failure is typically catastrophic
Flash memory (solid state drives or memory sticks)	10 – 13 years without use. May extend up to 100 years with active management.[153]	• Readily available • Convenient • Small/Portable	• Current size limitation is 8 GB per unit. • Loss is inevitable without active management.
Optical media (ROM)	20 – 12,000 years.[153]	• Readily available • Requires little technical knowledge • Potentially lengthy life expectancy. • Permanent (once written it is only readable). • Multiple densities available	• Requires dedicated drive technology for reading and writing. • Drive technology may become obsolete. • Need for secure physical storage location in which to retain media. • Limited unit storage size.
Optical media (recordable)	Light and heat dependent but can be as low as a few hours in direct sunlight[154]	• Readily available • Requires little technical knowledge • Multiple formats available • Rewritable (non-permanent)	• Need for secure and dark physical storage location in which to retain media. • Requires dedicated drive technology for reading and writing. • Drive technology may become obsolete. • Limited unit storage size.

1) Is the file format a standard base coding format? (e.g. UNICODE, ASCII)
   a. Yes – This file format is low risk.
   b. No – Continue to Question 2.
2) Is the file format referenced in any searchable information resources?
3) Is there a known support end date for the file format?
   a. Yes – How many years until the support end date? (if within your long term storage needs, consider an alternate format).
4) How many years since this file format version was released?
   a. Are new versions available or on the horizon?
5) What is the primary rendering software needed for this file format? (i.e. what program is needed to read the files).
   a. Identified – is this software available to you?
6) Does the primary rendering software have critical hardware or software needs that might not be available in the future?
   a. Yes – what equipment/software is required and is it available to you?
7) Are there alternate rendering software solutions available? If so, what are they and what are their hardware/software requirements?
   a. Identified – how many of these are available to you with all their requirements?
8) Are there other means of providing safe and effective access? (e.g. custom coding, open source applications?)
9) What is the total number of access methods available for your chosen file format? Include all primary and alternate methods.
   a. 0 – Extremely high risk, consider your data as lost.
   b. 1 – High risk, consider alternate formats if possible
   c. 2 to 5 – Medium risk, ensure hardware/software requirements for access are documented and retained if possible.
   d. 5 or more – Low risk, you can proceed with implementation.

Data Migration

Data migration involves the process of moving data from one source to another where the structure of the data will change.[5] Data migration might be necessary if a registry platform becomes obsolete (either due to changes in software design or due to software discontinuation); if software cost becomes an issue (in the case of proprietary software platforms especially); or if additional functionality not planned in the initial registry design is required. Further data migration may be required if physical server characteristics or locations change; if data ownership requirements or personnel change; or to meet larger IT infrastructure expectations within the host organization. While it is possible to do data migration manually, the time investment will be considerable and efforts should be made to select software that can mediate an automated migration. For an automated migration to be successful and detailed data map correlating every data type from the old system into the new system must be created.[5] A plan for handling inconsistent data should be created at the outset and revised if any additional issues are raised during the migration process.[5] Following migration, quality assurance activities should be conducted to ensure that the data in the new system has been transferred as expected. Overall, a simple project management methodology made popular by W. Edwards Deming of "Plan-Do-Check-Act" (the Shewhart Cycle) can be a great approach for a data migration project. First, plan the data migration including required staffing and software/hardware resources; project timelines and any server down time that may occur. As previously mentioned, ensure that this plan features a detailed map of the data migration from the old system to the new system. Next perform a small test migration. Following the test, enact your quality assurance plan and evaluate if the desired results have been achieved. Once you have reviewed the test results, either revisit the plans and retest, or proceed with the full migration.

https://doi.org/10.1017/S0317167100017157 Published online by Cambridge University Press

Data Curation

Data curation is defined by Lord and MacDonald as "the activity of managing and promoting the use of data from its point of creation, to ensure it is fit for contemporary purpose, and available for discovery and reuse".[156] This activity may include simple data management activities, enriching or adding value to data, the sharing of data, and the preservation of data for a later use. Data curation is a critical activity for the creation and maintenance of successful disease registries. While this guideline cannot define an individual registry's curation plan, below are some key points to ensure creation of a complete data management plan. These key points are taken from a Data Management Plan checklist produced by the Digital Curation Centre.[157]

1) Data types – understand what types of data will be collected in the registry. Ensure that data are defined using a data dictionary.
2) Data formats – consider the format of each type of data (e.g. text, alphanumeric, date etc). List all the possible formats that will be collected across the registry.
3) Standards – likely partly outlined in the data dictionary but ensure that there is clear definition of what data will be accepted and rejected. Additionally document if data will be compared to other sources and any associated standards dictated by this relationship.
4) Capture methods – document all methods of data capture and data flow into the data repository (database).
5) Data output – consider what content is being created by the project and document this. For example does the project simply produce raw data for further use or does the project produce derived data.[158]
6) Storage - what storage space is required for the data output? See Storage Considerations in this Guideline for more information.
7) File formats – what file formats will be used and why? Ensure that you document your analysis of file format considerations and risks in the data management plan. See Storage Considerations in this Guideline for more information.
8) Future uses – consider what the future uses or reuse of the data output and/or original data might be. What will be required to ensure these future uses/reuse can occur.
9) Sharing – ensure that consideration of whom might share the data and all associated ethical, legal and logistical issues are outlined and addressed.
10) Access – who will have access to the data and what are the access controls?
11) Existing data – are there existing data that are required or beneficial to the project. What constraints or considerations are present as a result? Is new data production actually needed? What is the value of the new data? What access is required to obtain existing datasets?
12) Data quality – what is your plan for data quality assurance and control?
13) Documentation – what documentation is required to ensure that data make sense in isolation? Consider that the context required may be stored with the data itself using metadata.
14) Metadata – if metadata will be included ensure you have considered how they will be created, maintained and stored.
15) Intellectual Property – ensure that the ethical and legal considerations associated with existing data and new data have been considered and addressed. See the Ethics & Privacy section of this guideline for more information.
16) Accountability – who is responsible for the data and who are the delegates of this authority if applicable? How is accountability assigned (e.g. legislation; institutional policy)? How will accountability be transitioned if required?

Data Management Plan

A data curation document will be part of a larger data management plan. The data management plan will include additional aspects such as:

- Who manages the data?
- Where, how and when will data be backed up?
- What mechanisms are in place for error tracking and change logging?
- Who is responsible for addressing changes, errors and trouble? What is the process for addressing changes, errors and trouble?
- What security systems are in place to protect the data?
- What is the process if there is a security breach?

For assistance creating a comprehensive data management plan, consider utilizing the DMPTool found at https://dmp.cdlib.org/.[159]

Recommendations

✓ In the context of applicable Canadian legislation consider the following items with respect to data storage:

- o Server Model
- o Physical Location of Servers and Access
- o Network Location of Servers
- o User Access levels and permissions
- o Hardware and software security controls

✓ Consider the complexity of your storage needs and the required personnel and software resources to maintain them.

- o Maximize organizational assets such as existing software licenses or discounts.
- o Wherever possible utilize open source software to minimize development and ongoing costs.
- o Document and plan your development timeline.
- o Ensure you have planned for adequate storage space and database size/functionality. Assess required computer hardware to facilitate desired access times; registry stability and needed reporting capabilities.

✓ Choose multiple data storage formats for short and long-term data backup. Ensure backup plans meet necessary legislation and policy expectations. Document data backup procedures and schedule in the data management plan.

✓ Assess file format storage risk.

✓ Create data curation and data management plans.

https://doi.org/10.1017/S0317167100017157 Published online by Cambridge University Press