ACKNOWLEDGEMENTS

The author wishes to acknowledge the following individuals for their detailed input, expertise and contributions to this report:

- Mahmoud Azimaee (Institute for Clinical Evaluative Sciences)
- Jack Rach (Manitoba Centre for Health Policy)
- Mark Smith (Manitoba Centre for Health Policy)
- Ruth-Ann Soodeen (Manitoba Centre for Health Policy)
- Ken Turner (Manitoba Centre for Health Policy)

How to cite this report:

Bond R. Data Documentation Framework. Winnipeg, MB. Manitoba Centre for Health Policy, October 2012.
# TABLE OF CONTENTS

Acknowledgements ........................................................................................................... 1
Overview ............................................................................................................................. 3

1. Background and Rationale .............................................................................................. 4

2. Documentation and Data Management ......................................................................... 4
   2.1. Data Acquisition (Steps 1-3) .................................................................................. 5
   2.2. Data Quality Assurance (Step 4) ............................................................................ 6
   2.3. Metadata Documentation (Step 5) ......................................................................... 7
   2.4. Data Release/Dissemination (Step 6) ..................................................................... 7

3. Data Documentation Framework ..................................................................................... 7
   3.1. Framework Scope ..................................................................................................... 8
      1) Documentation .................................................................................................... 8
      2) Audience/User Base ............................................................................................ 8
   3.2. File Management ..................................................................................................... 9
      1) Storage and Retrieval .......................................................................................... 9
      2) Security ............................................................................................................... 10
   3.3. Document Quality Indicators .................................................................................. 10
      1) Style/Format: Does this look like an MCHP document? ....................................... 10
      2) Provenance: Can I trust the contents of this document? ...................................... 11
      3) Evaluation: Is this document useful to me? ......................................................... 13
   3.4. Document Release .................................................................................................... 14

4. Managing the Framework ............................................................................................... 14

5. Next Steps .................................................................................................................... 16

References .......................................................................................................................... 17
A Data Documentation Framework was created by the Data Management Work Unit at the Manitoba Centre for Health Policy (MCHP) to guide members of this unit in developing policies and procedures for managing data documentation. It identifies elements of good documentation practices and desirable characteristics for the documents arising out of the data management process. Four attributes of good documentation suggested by the literature are included in the framework: a clearly-defined scope, optimal file management practices, explicit document quality indicators, and a description of the dissemination/release process.

1. Scope describes the documentation and audience/user base to include in the framework.
2. File management identifies indicators for storage/retrieval and security practices for handling files.
3. Document quality includes indicators addressing the following key questions:
   - Style/format – Does this look like an MCHP document?
   - Provenance – Can I trust the contents of this document?
   - Evaluation – Is this document useful to me?
4. The last attribute identifies distribution and communication as the main elements in the dissemination/release process.

This document concludes by itemizing the tasks required to ensure good documentation practices take place within an organization, and describes how a dedicated documentation manager might coordinate these tasks through the use and application of the Data Documentation Framework.
This report introduces a framework for managing the documentation typically created and/or compiled by the Data Management Work Unit throughout the data management process at the Manitoba Centre for Health Policy (MCHP). This process is described, illustrating the kinds of documentation normally associated with the acquisition and management of databases. While all work units within MCHP tend to follow certain documentation practices as part of the larger organization, it was felt that the Data Management unit and – by extension – MCHP might benefit from a framework proposing several explicit indicators, as well as suggestions for managing the framework.

1. BACKGROUND AND RATIONALE

Documentation is created and used in a wide range of circumstances in the Data Management Work Unit at MCHP, from brief descriptive summaries of databases for an external website audience to legal forms ensuring government requirements are followed for approving and monitoring access to these databases. Documentation is foundational to the production of research at MCHP; however, there are few regulatory or legislative requirements directly governing the quality of various aspects of documentation. A number of resources are available spelling out measures for improving various elements of documentation, but few are comprehensive and systematic.

Repercussions for poor quality documentation can be as significant as the consequences associated with the use of poor quality data (M. Azimaee, 2011). In fact, one clinical researcher stated that standard operating procedures for good documentation should always be developed since the quality of documentation "can make or break the study at a given site" (Bargaje, 2011). Not only poor quality data, but also poor quality documentation can result in drawing false conclusions and implementing inappropriate policies and procedures. Outdated, badly written or annotated documentation, combined with inconsistent storage practices, can result in huge costs to an organization.

Several studies highlight the importance of having easy-to-find, easy-to-read, approved, current, accurate documentation. Some estimates (Laserfiche, 2007) suggest that the typical worker takes about 12 minutes to process a paper document, with 9 minutes taken up with searching for, retrieving, and re-filing: only three minutes are spent actually using the information. As well, it has been found that employees can spend 20% of their day hunting for information and that 50% of the time they cannot find what they need. Auditors may also pay special attention to document control practices and quality management systems (Dawson, 2006).

As a research unit within the University of Manitoba, the Manitoba Centre for Health Policy is subject to university policies and procedures and, as a trustee for the Population Health Data Repository, is subject to provincial government policies and procedures. Not only must legal expectations for documentation from organizations outside MCHP be met, the documentation must support and facilitate the production of high-quality research.

To address these issues, a Data Documentation Framework was created by the Data Management Work Unit at MCHP to identify elements of good documentation practice and desirable characteristics for the documents arising out of the data management process.

2. DOCUMENTATION AND DATA MANAGEMENT
A range of documentation is compiled and/or developed throughout the six-step process of managing data at the Manitoba Centre for Health Policy (MCHP). For each step in the Data Management framework (Table 1), examples of documentation are shown, beginning with material received/processed at the early stage of negotiating with the source agency (Step 1) to the final step of releasing both data and documentation to potential users of the database (Step 6).

The Data Management Work Unit recently formalized this process with particular attention to the areas of data quality evaluation, documentation and the release and distribution of information concerning new acquisitions (Cadham deliverable, 2012). This process is consistent with standards and practices suggested in similar initiatives and follows recommendations developed by other organizations who develop and maintain repositories of anonymized personal health information for research purposes (Lyman et al., 2008; Daas et al., 2008; Holman et al., 1999).

Approximately 60 databases are currently housed in the Repository, with another 16 databases expected to be acquired over the next few years. The MCHP Data Management Work Unit ensures that this rigorous 6-step process is applied to each Repository database.

### 2.1. DATA ACQUISITION (STEPS 1-3)

Before databases are acquired at MCHP, the first piece of documentation – a data sharing agreement - must be completed, defining policies and practices with regard to data confidentiality, privacy, legislative and regulatory requirements, data transfer, and ongoing use of data for research purposes. Once this is established, the data management analyst assigned to the given database works with the source agency to acquire all available documentation related to the database. This documentation is helpful for developing a formal data request to the agency.

After the request is formulated and the data are received, an MCHP data management analyst ensures basic database documentation is as complete as possible (Step 2). This may include filling in missing gaps in data documentation such as creating data models.
For the third step in the data management process, an MCHP data analyst uses SAS® software to prepare and install the data in the SAS Scalable Performance Data Server (SPDS). The installation process includes conversion of files to the appropriate format and application of standard naming/formatting conventions.

### 2.2. DATA QUALITY ASSURANCE (STEP 4)

Both quantitative and qualitative methods are used to assess data quality, and tools such as SAS macros are used to automate the process wherever possible. Results are summarized in the data quality reports now routinely generated at MCHP for each database to summarize its fitness for research use. Since the degree to which a database can be investigated and evaluated is determined by the data access approvals received by the researcher, the Data Quality Assurance framework (Figure 1) distinguishes between database-specific (Repository-installed) and project-specific quality assessment processes.

![Data Quality Assurance Diagram](image)

While data quality reports already may be available from the agency or from previous research and contain helpful information about the data, they may not contain the desired level of comprehensiveness. Given the widespread use of data housed at MCHP to produce top-quality, reliable, and trustworthy research, it is critically important to ensure systematic, standardized evaluations of all databases housed in the Repository are made available.

The Data Quality Assurance Framework was developed by MCHP after a thorough review of existing frameworks from similar organizations such as the Canadian Institute for Health Information and Public Health Agency of Canada. It encompasses various dimensions of quality such as accuracy, validity, timeliness, and interpretability – all of which can be measured by means of one or more indicators. MCHP data quality reports thus provide important user information, in both table and graph formats, such as number of missing values, proportion of valid/invalid values, and minimum, maximum, and mean values for numeric fields. This provides the user with a “head start” in the always-needed preliminary work of checking and cleaning the data to prepare it for analysis.
2.3. METADATA DOCUMENTATION (STEP 5)

This step ensures the documentation compiled throughout Steps 1 to 4 is accessible through a single point of access. While “metadata” is often defined as “data about data”, the MCHP Metadata Repository more specifically refers to the documentation associated with the various databases housed in the MCHP Data Repository. Designed to be a one-stop location for accessing a comprehensive range of documentation, tabbed pages provide a consistent structure to the delivery of content. By clicking on any given database, the user is taken to a page with the following standard tabs: 1) Overview, 2) Data Model, 3) Data Quality, 4) Data Dictionary, 5) Documents and Reports, and 6) a blog area for discussions about the database.

2.4. DATA RELEASE/DISSEMINATION (STEP 6)

Once Steps 1 through 5 have taken place, the data and data documentation are released for use by researchers and data analysts/programmers who have obtained the appropriate access permissions. This release may be announced internally via an informal process (e.g., by notifying data analysts/programmers at MCHP that the new data and documentation are available for use) and/or through a more formal process (e.g., presentations to internal analysts and researchers). At around the same time, a descriptive overview of the data is published to the external MCHP website and links are included to the new material from the list maintained for the holdings in the MCHP Data Repository.

3. DATA DOCUMENTATION FRAMEWORK

A review of the literature suggests attributes of good documentation should include a clearly-defined scope, optimal file management practices, explicit document quality indicators, and a description of the dissemination/release process. In addition, for MCHP, data documentation practices must recognize that the organization operates within the context of the University of Manitoba and government legislative requirements.

The Data Documentation Framework is depicted as a circle (Figure 2), reflecting its iterative nature. Even though determining scope can be seen as a first step, it should be constantly evaluated to make sure it is continuing to meet the needs of the organization, accommodating changes in: a) the audience/user base and b) the kinds of documentation that should be included in the framework. Step 2 – File Management – refers to the physical handling of files: a) their storage and retrieval, and b) their security. Step 3 – Document Quality Control – focuses on what is contained in the file: a) style/formatting of contents, b) trustworthiness of the contents, or its provenance, c) how the contents are used, and d) processes to evaluate the quality of the document. The last step – Document Release – refers to dissemination of documentation: a) communicating that the information is available and b) distributing the contents or location of the document(s).
3.1. FRAMEWORK SCOPE

Framework scope consists of two main areas, basically defining what (documentation) and who (audience/user base) are included.

1) DOCUMENTATION

“Documentation” can include media of all types (electronic files, hard-copy files, tapes, etc.) and refers to both internally- and externally-produced materials. External documents might include items such as national/international standards, manufacturer’s manuals and equipment software, all of which should be included in a review process. They may be subjected to some form of control, for example, by “documenting, as appropriate, the title, document number (if any), most recent and/or approved revision/edition number, and year of publication” (US EPA, 2011).

At MCHP, the scope of the Data Documentation framework was defined to include all materials related to databases, including those developed at MCHP and those coming into the organization, and consisting of all documents managed by the Data Management Work Unit.

2) AUDIENCE/USER BASE

The audience is the target group for whom the documentation is intended. Audience is a critical element for all documentation at MCHP, since confidentiality requirements dictate what is permissible to release and to whom. Internally, the audience to which the Data Documentation framework applies encompasses a wide range of users - from data management analysts/programmers to support staff to researchers - who may be working within the organization or from remote access sites. Externally, another wide and multifaceted group

Figure 2 - MCHP Data Documentation Framework
of potential users (e.g., students, researchers, policy analysts) are able to access non-confidential information at the MCHP website on the WWW.

3.2. FILE MANAGEMENT

File management refers to how files - whether electronic or hard-copy - are handled in terms of storage/retrieval and security practices. MCHP has strict protocols governing the handling of data files, with practices currently evolving for handling associated documentation. The Microsoft SharePoint Content Management System (CMS) instituted in 2006 at MCHP has helped to facilitate optimum file handling techniques however the former system of using standard Windows directories continues to be used as well. This has resulted in some confusion concerning how files should be handled in the Data Management Work Unit, and also more generally in MCHP.

1) STORAGE AND RETRIEVAL

Policies and processes for storage and retrieval (Dawson, 2006; USA EPA, 2011) should include those describing: a) retrieval mechanisms, b) file naming practices, c) location, d) retention/destruction schedule, and e) back up schedule.

a) Retrieval mechanisms – Microsoft SharePoint has a powerful search engine, which searches not only material stored within the MCHP SharePoint portal but also material created and stored elsewhere in other locations on the server. The Metadata Repository set up by the Data Management Work Unit (Step 5 of the Data Management process) offers another option for finding information internally at MCHP.

b) File naming practices – MCHP has general guidelines in place on managing files in directories, including naming practices for folders and suggestions for annotating and referencing practices. Improvements to current practices have now been made possible with the introduction of SharePoint. Enabling the versioning feature of SharePoint, for example, permits maintaining the same name for a document and allowing earlier versions (also with the same name) to be retrieved as needed. These earlier versions are dated and can be annotated to facilitate selection by the user.

c) Location – Dawson (2006) notes the importance of establishing designated, controlled locations for documents. The easier it is to access such copies, the more likely they are to be used (as opposed to uncontrolled copies). One organization stores official copies of all documentation authorized for use on their LAN, and all other copies are considered unofficial (USA EPA, 2011). Within MCHP, this practice has been adopted for databases, which are stored in an official location once processing has been completed by the Data Management team.

d) Retention/destruction schedule – A good search engine enables users to find a required document. However, when a search returns a list of multiple files appearing to be variations of the same document, it may not be possible to determine the official version unless good documentation practices have been followed for the content. Dawson (2006) indicates that segregation or disposal is usually required to prevent unintended use of obsolete documents, adding that “any obsolete documents that are kept for
reference or other purposes must be clearly identified through markings, separate storage areas, or other means”.

e) **Back-up schedule** – MCHP has a strict policy on backing up all files – database and documentation – and this takes place on a regular schedule. Offsite archival and back-up systems are also in place.

2) **SECURITY**

Policies on access, security, and confidentiality are in place at MCHP for handling of data and associated documentation, and are regularly reviewed in conjunction with legislative requirements. The Windows server and SharePoint permit a wide range of permission settings, offering the capability to restrict access for a single document, for designated individuals, or for groups of individuals.

3.3. **DOCUMENT QUALITY INDICATORS**

While it is possible to address many components of data quality using the explicit indicators automatically generated by macros in the MCHP Data Quality Framework, assessing the quality of documentation requires a significantly greater subjective component in producing useful metrics. Checks for certain kinds of invalid responses in data (e.g., February 31) can be built into macros, but different techniques are needed to measure indicators such as validity in assessing the quality of a document. The MCHP Data Quality Framework has referred to this need for quality documentation for all databases as **Interpretability**.

We propose that (in addition to standard spelling/grammar checks) quality control for documentation address three key questions:

1) **Style/format** – Does this look like an MCHP document?
2) **Provenance** – Can I trust the contents of this document?
3) **Evaluation** - Is this document useful to me?

**1) STYLE/FORMAT: DOES THIS LOOK LIKE AN MCHP DOCUMENT?**

The look of documents prepared by MCHP for public use is determined to a certain extent by the University of Manitoba Visual Identity Guide; this applies primarily for material available on the public website and to Powerpoint presentations. The Visual Identity guide provides extensive detail on the use of logos, logo placement, typography, and other visual style elements. For other public release documents such as deliverable reports, style guides were created at MCHP to ensure consistent application of formatting standards (e.g., desirable terminology, reference/citation standards). For documents prepared for restricted use or for in-house use, visual identity guidelines are not currently applied at MCHP.

Baldwin (2005) stresses the desirability, where possible, to have in-house organizational documents resembling each other in layout, tone, or simple look and feel. Standards for the look of in-house MCHP documents might include those at a very broad level, covering the use of such items as logos, headers, footers (e.g., style of page numbering, date format), and references. Individual work units may go further in
developing guidelines specific to the work they produce. Within the Data Management Work Unit, for example, guidelines are under development to ensure a consistent look and feel for data dictionaries.

2) PROVENANCE: CAN I TRUST THE CONTENTS OF THIS DOCUMENT?

In the art world provenance research is important for determining authenticity, for assigning value, and for determining ownership (International Foundation for Art Research, 2012). Provenance in the open and inclusive environment of the web essentially refers to authenticity, providing a critical foundation for making determinations about whether the information can be trusted (XG Provenance Wiki). From an administrative perspective, trustworthiness also has implications for operational continuity and historical legacy/research.

The concept of provenance can be applied to documentation generally through two key components: a) Document Profile (metadata) and b) Document Content (Barry, 1994). Both help to provide strong evidence of trustworthiness, including assurances of authenticity and competency. With this kind of information, the user is able to make an informed judgment as to the trustworthiness of a document by checking its profile, or metadata, and by having some knowledge of guidelines followed by creators and revisers in working with documents.

a) Document Profile: Metadata

To help in establishing the provenance of any given document, certain basic kinds of information are helpful to routinely include for all documents (Table 2). Metadata items may be consistently applied across all documents or their application might vary, depending on whether the document is intended for external or internal use. It is important not only to include metadata, but also to include it in a standard way—e.g., on the first page of every document. (Even after a document is destroyed, its metadata can be kept for administrative purposes.)

Metadata helps make it clear to the user who created the document, when, and why (Barry, 1994; Liepa, 2004; Millman, 2007; US EPA, 2011). Many documents at MCHP contain information for some of the items listed in Table 2, but no guidelines are in place to ensure standard collecting of metadata across the organization. To facilitate this process, the portal content management system for the internal network offers excellent tracking and monitoring mechanisms for metadata maintenance as part of the document review process. They include:

- **Access permissions** – range from viewing only to full edit capabilities for any given document. These can be set by user or by user group.
- **Check-out capabilities** – permit collaboration by multiple users by blocking access if work is being performed on the document.
- **Workflow routing** – permits assigning reviewers and monitoring the progress of the document as it goes through a

<table>
<thead>
<tr>
<th>Table 2. Document Profile: Metadata</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Title – might also include a document number, if a document numbering system is in place.</td>
</tr>
<tr>
<td>• File name – using standard naming conventions as per file management standards, if in place.</td>
</tr>
<tr>
<td>• Author – original author. This information may also let users know who to contact for corrections or updates.</td>
</tr>
<tr>
<td>• Department/work unit – if applicable.</td>
</tr>
<tr>
<td>• Document type – noting categories like manual, procedure, policy, or form may be sufficient for some documents while for others, it may be important to note, for example, that it is one of a series.</td>
</tr>
<tr>
<td>• Document purpose – might also include the target audience.</td>
</tr>
<tr>
<td>• Effective date – denotes the date from which the document is in force. It indicates that the final version has been approved. This might include authorizing signatures.</td>
</tr>
<tr>
<td>• Security level – is important for ensuring that confidential documents are not accessible to a wider audience. Levels could include public/www, restricted to designated users, etc.</td>
</tr>
<tr>
<td>• Legal information – disclaimers, trademarks, copyright, etc. should be included as necessary.</td>
</tr>
<tr>
<td>• Document version/history (and/or revision number) – can range from draft to finalized for release to users. Displaying sign-off information confirms to users that the document is ready for use.</td>
</tr>
<tr>
<td>o If a revision – date of change, nature of change, and person(s) authorizing the change.</td>
</tr>
<tr>
<td>o If a full review – date of review and person who did the review, as well as nature of any changes.</td>
</tr>
</tbody>
</table>
specified workflow.

- **Versioning** – maintains historical versions of a document which can be easily accessed – all without having to change the name of the file. This feature is especially useful for documents which require periodic updating and/or where previous versions must be easily accessible.

Standard templates for metadata entry (and for document styles/formats) might also be set up for use in documents created within the portal environment.

### b) Document Content: Review and Approval Process

In addition to the information that can be obtained through the metadata of a document, a user can further assess its quality if information about the review and approval process for the document is available. Document review is the process “through which persons with subject matter knowledge contribute to the development of internal documents”; this process should include grammatical, editorial and technical assessment (US EPA, 2011), and applies to both the creation of new documents and the review of existing documents.

The process for document review should be spelled out for various categories of documentation (e.g., data dictionaries), including the appropriate use of templates and the vetting/signoff process for new documents. It should include a description of the error correction process and describe how continuing suitability and compliance with any applicable requirements are carried out (US EPA, 2011).

Context is also important. For example, if a document is one in a series of documents, this should be explained up front (Millman, 2007): “people work better, learn better, and will be more successful with your documentation if they know where and how it fits together”.

Reviewers involved in the process may include “future users of the document, managers responsible for the activity, workers in areas affected by the activity and other interested individuals” (Dawson, 2006). It is suggested that a review cycle be built into the document development procedure and that a record be kept of such reviews (Dawson, 2006).

For data dictionaries, for example, CIHI (2009) applies several standards, including evaluation of all data elements to determine their inclusion in the CIHI data dictionary and ensuring the latter conforms to dictionary standards, with consistent data attributes. For their Data Quality Framework, CIHI (2009) explicitly states that three types of documentation about a data holding’s data quality are required: data quality assessment report, data quality documentation for users, and metadata documentation.

The following are some examples of processes currently followed to ensure quality of documentation at MCHP:

- **Summary database descriptions** – are written using a standard template consisting of Database Name, Source Agency, Type, Purpose, Scope, Data Collection Method, Size, Data Level, Data Years, Data Highlights, Access Requirements, and References. The review process ensures the draft goes to researchers and analysts at MCHP who have worked with the data, as well as to all members of the Data Management Work Unit. Following a final review by the external data agency from whom the database was received, the database description is released by uploading it to the WWW on the MCHP external website.
• **Concept Dictionary** – contains detailed definitions related to methodologies for working with the data housed in the MCHP Data Repository. After initial drafting (usually by a data management analyst) - and following explicit guidelines for both structure and content - the concept is circulated to all individuals who have worked directly with the methodology, including researchers and/or programmers involved in its development.

• **Deliverable Report Content Review Checklist** – serves to illustrate one of the review and approval processes used in other MCHP work units. Providing instructions to team members about how to complete the checklist and how to process it, the list includes:
  - Acknowledgements – who should be included in various categories
  - Introduction - e.g., “Background is current, appropriate, and sufficient”
  - Methods – clearly and accurately describes e.g., “Specific data sources (e.g., databases, files) used”
  - Results – e.g., “Explanation for reading and interpreting graphs and tables is provided if necessary”
  - Discussion/Conclusion – e.g., “Connection made to background and study objectives”
  - References – e.g., “Citations in text match those in the Reference List”
  - Glossary – e.g., “Acronyms used in the report are listed at the beginning”

Review of all documentation should regularly take place to ensure ongoing, reliable levels of quality. It is recommended that an annual documentation review take place to “spot redundancies, documents no longer needed, and opportunities to consolidate” (Dawson, 2006). Realistically, an annual review may not be possible for all documents. One organization systematically reviews their internal documents at least once every four years (USA EPA, 2011).

To facilitate systematic review, it is helpful to maintain a master list of all existing documents (US EPA, 2011), or documents that should regularly be reviewed. To facilitate tracking and monitoring of a document during the document review process, systems like SharePoint can be used to automate this process. This is a feature that the Repository Access Work Unit at MCHP regularly uses to track the document approval process for obtaining access to the data repository. The document review schedule and process should be spelled out in a policy/procedure document.

---

3) **EVALUATION: IS THIS DOCUMENT USEFUL TO ME?**

Determining the audience for a document is key. Confidentiality rules governing the release of information at MCHP dictate, for example, whether the audience will be a very restricted internal user group or the world-wide audience of the internet. Level of expertise and familiarity with the topic is also important: highly technical writing is normally targeted to a small user base.

Evaluation approaches can draw from a mix of quantitative and qualitative information gathering to ensure methods appropriate for the milieu and the audience are selected.

• **Quantitative** – Frequency of use for documents located in a web-based environment like SharePoint or the internet can be determined automatically through methods such as counting the number of times a document is opened. Dawson (2006) also suggests tracking the number of document revisions due to information mistakes, indicating results will often reveal weaknesses in the review/proofreading process.
• **Qualitative** - Employees might be periodically surveyed regarding the usability of any given document. This method can help determine whether the documentation is serving its intended purpose. Blogs are another way to obtain user feedback, as are popup mini-surveys activated as the user accesses the document.

Reports may also be generated on the quality of any given document; for example, a summary indicator might be noted on the document itself, or a summary report might be generated to accompany a master list of documentation.

### 3.4. DOCUMENT RELEASE

For documentation to be useful, it must be made available in a timely fashion. CIHI (2009) measures currency of data quality documentation, for example, by its availability at the time of data or report release. Once a document is finalized, its distribution details should be spelled out and an announcement should be communicated to relevant user groups.

- **Distribution** – Several factors come into play when releasing a document at MCHP. Its security level will determine which user groups will be able to access the document. This will also determine where the finalized document will be located. For example, if it is to be a general public release, it may be stored in a WWW folder within a SharePoint folder for uploading to the University of Manitoba server, or designated suitable for release in other ways.

- **Communication** – MCHP regularly sends out newsletters with announcements of new deliverable reports, new databases, and new concepts to external organizations and individuals on a regularly-maintained distribution list. These newsletters include links to the new material. Internally, a system is being developed to ensure that information is conveyed internally about new or revised documentation restricted to internal users.

### 4. MANAGING THE FRAMEWORK

Because of the importance of documentation control, companies often invest heavily in “dedicated staff, detailed procedures and specialized software programs” (Dawson, 2006). One example is the Document Control Coordinator (DCC) employed at the Science and Ecosystem Support Division (SESD) of the United States Environmental Protection Agency (2011). This is a quality management position appointed to ensure “the most recent version of the procedure is placed on the SESD LAN and for maintaining records of review conducted prior to its issuance” (US EPA, 2011). “Subject Matter Experts” deemed competent, experienced and knowledgeable by management contribute to the content of the document.

To help determine if/how the proposed MCHP Data Documentation Framework might be managed, the steps in the framework were compared with the tasks carried out by the Document Control Coordinator at the EPA (Table 3).
<table>
<thead>
<tr>
<th>MCHP Documentation Framework</th>
<th>Tasks carried out by the Document Control Coordinator at US EPA</th>
</tr>
</thead>
</table>
| **Step 1. Framework scope (also includes elements of file management and quality control)** | - Maintain a master list of all documents, both internal and controlled external documents. The latter are reviewed by the DCC once a year; this process includes consulting with the source agency to determine if updates are available. If yes, the obsolete copy is reviewed and the update is labeled as such. External documents may include items such as national/international standards, manuals and directives from governing agencies.  
- Maintain policy/procedure documents spelling out the process for development, review, authorization, control, and distribution of controlled documents. |
| **Step 2. File management** | - Mark obsolete copies with a watermark and/or notes date of archival. |
| **Step 3. Document quality control** | - Maintain a document review schedule which follows the US EPA 8-step process ranging from notifying the author of the document of the need for review to submitting the document to management for final review and signatures.  
- Coordinate a final format check, authorization, and distribution of internal documents after they have completed final review. |
| **Step 4. Document release** | - Notify all personnel of document updates. This includes notification of documentation that has been archived/replaced, advising disposal of old copies to preclude unintended use. |

Other tasks might include:

- Educate staff on the need for quality processes through forums such as workshops, including training in how to document to the standards prescribed by the framework.
- Communicate with other work units within the organization to stay on top of other possible quality assurance initiatives.
- Regularly evaluate the framework in terms of its usability, i.e., whether it is serving the intended purpose.

Extensive processing is required throughout the steps of acquiring, installing, and documenting each of the databases housed in the MCHP Repository. Ensuring production of high-quality documentation adds another, significant layer to this processing. As well, the Repository continues to grow steadily, with 16 new databases expected to be acquired over the next several years through the CFI Initiative. Yet others will be added during this period to meet the needs of specific deliverables. All of this information needs to be carefully documented so that it can continue to be of use to MCHP in the future.

The tasks identified by the US EPA for a quality management position currently tend to take place on an ad hoc basis within MCHP, and within the Data Management Work Unit. A systematic approach, one supported by management, is needed to coordinate and ensure quality control of documents at MCHP through completion of such tasks, which likely will require a dedicated FTE.
The Data Management Work Unit at MCHP currently consists of six individuals reporting to the Associate Director, Repository, including three full-time data management analysts, one .8 FTE data management analyst, a full-time lead data acquisition officer, and a data documentation manager (.8 FTE). With data management analysts/programmers increasingly contributing to creating the necessary data documentation and with increasing attention to ensuring all databases are accompanied by appropriate documentation, the role of Data Documentation Manager at MCHP is uniquely positioned to develop and evolve into one dedicated to the management of the MCHP Data Documentation Framework.

5. NEXT STEPS

- Identify and summarize current strengths and weaknesses re MCHP documentation.
- Action plan for improvement (prioritization and implementation) - to maintain strengths and work toward resolving weaknesses.
- More details for each piece of the framework as relevant/necessary (e.g., documentation quality checklist suggested by Mahmoud; Table 2 of this report describing metadata that can be used as a standard template for documents)
- Sharing and developing the framework: collaboration and networking
REFERENCES


