HIFIS

Homeless Individuals and Families Information System

Data

Management

Guide

Version 1.0

Table of Contents

[About this Guide 2](#_Toc64293644)

[Background 2](#_Toc64293645)

[SECTION 1: The Value of and the Importance of Managing Data 3](#_Toc64293646)

[1.1. Why data matters 3](#_Toc64293647)

[1.2. Demonstrating the value of data in a community 3](#_Toc64293648)

[SECTION 2: The Role of a Data Steward 6](#_Toc64293649)

[2.1. Understand the needs of data users 7](#_Toc64293650)

[2.2. Develop and enforce standard data management processes 7](#_Toc64293651)

[2.3. Manage accessibility and security of data 7](#_Toc64293652)

[2.4. Monitor overall data quality and resolve common data quality issues 8](#_Toc64293653)

[SECTION 3: Introduction to the Data Lifecycle 9](#_Toc64293654)

[3.1. Planning 10](#_Toc64293655)

[3.2. Acquisition 11](#_Toc64293656)

[3.3. Processing 13](#_Toc64293657)

[3.4. Analysis and Reporting 14](#_Toc64293658)

[3.5. Preservation and Storage 15](#_Toc64293659)

[3.6. Documenting with Protocols 17](#_Toc64293660)

[SECTION 4: Ensuring High Quality Data 18](#_Toc64293661)

[4.1. Ensure relevancy 18](#_Toc64293662)

[4.2. Standards for consistency 19](#_Toc64293663)

[4.3. Do data quality checks 19](#_Toc64293664)

[4.4. Invest in data literacy 19](#_Toc64293665)

[4.5. Support informed use of the dataset 20](#_Toc64293666)

[Conclusion 21](#_Toc64293667)

# About this Guide

This guide is intended to help HIFIS Leads and Coordinated Access Leads to value data as a tool to support effective service delivery and decision-making within the housing and homelessness response system.

# Background

Under Reaching Home, communities are accommodated to develop and deliver data-driven system plans with clear outcomes. This outcomes-based approach gives communities flexibility to address local priorities by investing in a range of interventions, including homelessness prevention initiatives as well as programming designed to meet the housing, income, and support needs of Canadians.

A key component of Reaching Home is the community-wide implementation of Coordinated Access. Coordinated Access requires strong local governance and collaboration, resulting in standardized workflow processes used consistently across the system. Responding to homelessness in this highly coordinated way helps communities to achieve their predetermined community-level outcomes by aligning local policies and service delivery approaches.

As the Reaching Home Homelessness Management Information System (HMIS), the Homeless Individuals and Families Information System (HIFIS) is key to the implementation of an effective, high quality Coordinated Access system. HIFIS gives service providers and local leaders access to real-time data about the current demand for services and capacity to meet this need. Communities use HIFIS data to support their local system planning efforts, enabling data-driven decision-making about performance management, progress with reaching community-level outcomes, and investment planning to close gaps in the system for those who are currently underserved.

# SECTION 1: The Value of and the Importance of Managing Data

## 1.1. Why data matters

Having access to high quality data, at the right time, can support organizations and communities in many ways. For example, data can clarify existing issues, identify new ones, and provide insight on how to address them.

High quality data is the product of effective data management practices. Commitment to these practices helps to ensure that the time and effort invested in data collection can support informed decision-making. Data can be useful for service delivery, decision-making within an organization, and for the community as a whole.

To achieve this goal, organizations and communities will need to establish new processes or formalize existing ones, and build internal capacity related to data stewardship.

This guide provides advice for strengthening data management practices by:

1. Introducing the role of a data steward;
2. Describing the concept of the data lifecycle; and,
3. Identifying common sources of data quality issues and approaches for resolving them.

While data may be perceived to be a technical and complex subject, this guide will illustrate how data management activities can be supported by effective strategies that improve data coordination. By taking incremental steps over time, practising good data management is something that everyone can do.

## 1.2. Demonstrating the value of data in a community

The fictitious community of Grandview is used throughout the guide as a setting to demonstrate and give context to different data management concepts. Grandview is a medium-sized community that is in the process of implementing Coordinated Access and is using HIFIS (Homeless Individuals and Families Information System) to manage their data.

The value of data is highlighted across the following brief examples:

|  |
| --- |
| **Example 1: Strategic Community Investments** |
| The community of Grandview is seeking to reduce inflows to homelessness by 25% for youth. They dedicate new funding to provide supports for youth at high risk of homelessness. They decided to dedicate this new funding to support after looking at evidence from prevention and diversion service providers, Porchlight. Porchlight reviewed their HIFIS data and identified a significant number of youths that were calling in more than twice a week to get help mediating issues in their family home, and sometimes to get support to stay with a relative for a few days a month so that they could avoid a shelter stay. The data that Porchlight collected helped Grandview prioritize 10 youths for support, which prevented them from becoming homeless.  |

|  |
| --- |
| **Example 2: Advancing Understanding of Homelessness in Canada** |
| In exchange for using HIFIS, communities provide the Government of Canada with 38 non-directly identifiable export fields. The export fields provide the Government of Canada with the data to inform policy, analysis, research, and evaluation purposes, which helps to advance the understanding of homelessness in Canada. In addition, the various types of data collected through HIFIS are used to create reports, such as the [National Shelter Study](http://publications.gc.ca/collections/collection_2017/edsc-esdc/Em12-17-2017-eng.pdf). This report contains findings on shelter use across the country, and the demographic characteristics of clients that access emergency shelters. This information allows INFC to better understand homelessness trends, and inform policy development for the homeless-serving sector.  |

|  |
| --- |
| **Example 3: Data for Community Planning** |
| The homeless shelter, Grandview’s Men’s Shelter, provides a monthly summary from data collected in HIFIS about where clients were before accessing the shelter. Grandview prepares the summary for local hospital, correctional facilities, and child welfare systems. The summary is used by the homelessness response system in Grandview to set targets to reduce the number of people discharged to homelessness from these places and helps to test solutions to reach the targets.  |

|  |
| --- |
| **Example 4: Impact of poor data management**  |
| In a Coordinated Access system, multiple client records could be the reason for a client not receiving the supports they require. Stacy, the Coordinated Access Lead in her community, is working to merge several duplicate client records to ensure her Unique Identifier List is complete. Each client has up to five different records from different service providers in the community. This process is challenging as some basic client data, such as date of birth, is conflicting in different entries for the same client. Furthermore, Stacy has no way of knowing how many duplicates exist for each client, particularly as some client’s names are input differently across service providers (e.g., Kate Hudson, Katie Hudson, Kat Hudson). To solve these problems and verify which data is correct for each client record, Stacy pulls the data from HIFIS and discusses with her case managers at the next team meeting. For the ones that can’t be resolved, the team agrees the only option is to contact the service providers where each client is being served and ask their staff to communicate with each client to personally verify conflicting data. This process takes several weeks and demands additional work from multiple team members.  |

# SECTION 2: The Role of a Data Steward

The data steward role is the “caretaker” of a dataset and serves as a local champion for adopting a data-driven culture in the community. Data stewards are responsible for ensuring that the data within their care is effectively managed throughout the data lifecycle (explained in Section 3).

To inform their work, they will need to understand the data needs of various stakeholders, including providers that contribute to the dataset and those that will use the data from it. Data stewards will also need to have a deep understanding of how the dataset can meet these various needs. Data stewards will use this knowledge to inform the development of effective data management practices. They will also play a lead role in promoting these practices with people who engage with the dataset across the community. In addition, data stewards are responsible for establishing clear communication channels for questions or concerns regarding the dataset, managing accessibility and data security, monitoring overall data quality and resolving common data quality issues. Ultimately, data stewards are accountable for data quality. Their efforts to increase data quality and users trust in the dataset help to strengthen the value of the data.

While the scope of responsibilities associated with data stewardship may vary, there are some common elements that organizations and communities should take into consideration:

* The role and responsibilities of the data steward should be documented as part of broader data governance in an organization or community.
* Multiple data stewards within a single organization or community will need to collaborate with each other to coordinate data management activities. A delegation matrix can be used to define shared responsibilities. For example, one staff member might be the steward for client records, while another might be the steward for extracting data and processing, while a third staff member could generate the monthly reports.
* The role of data steward should be filled by individuals with strong technical, communication, and leadership skills. For example, they should:
* Feel confident working with data and have previous experience doing so.
* Have a high level of subject matter expertise, allowing for data to be managed in a way that fits the sector
* Be able to collaborate with others and reinforce the use of policies and protocols
* Be able to train and educate staff

## 2.1. Understand the needs of data users

The data steward should work to understand the potential application(s) of the dataset under their care and manage the data accordingly. This role is particularly relevant in situations where the dataset in question is accessible to a variety of service providers, where the data may be used in different ways by different organizations. In these circumstances, a data steward may need to convene data users for regular meetings where they can share their needs and preferences. Data stewards could also seek to understand the needs of data users through other communication channels, such as online surveys or personal interviews.

Where data users have competing needs and interests, the role of the data steward will be to support prioritization and tailor data management activities to meet the unique needs of data users as necessary.

## 2.2. Develop and enforce standard data management processes

Based on the needs of data users, the data steward will be responsible for standard data management processes and activities. This could include developing tools, such as:

* A data dictionary;
* A data entry guide;
* Data integrity tools; and,
* Training curriculum.

These tools are supported by additional documents specific to communications, for example those that clarify the communication channels for questions or concerns regarding the dataset, such as a Frequently Asked Questions sheet. The development of various standards may be a collaborative effort with significant contributions from stakeholders, including data users and staff who are collecting and inputting the data. In this case, the data steward may be responsible for convening stakeholders, organizing working groups, and facilitating discussions. Developing standards may also be done in a more “top down” manner, conducted primarily by the data steward with input from stakeholders.

While standardized data management processes can improve data quality, they are only effective if used consistently by all users. As such, data stewards are also responsible for promoting the use of agreed-upon processes by reinforcing key messages, providing training and other communication tools.

## 2.3. Manage accessibility and security of data

Data should be accessible only to those who have the required training and authority to use it. The data steward may have to work in collaboration with senior managers and legal consultants or experts to determine specific criteria for accessing data. Data stewards will need to ensure that local policies and protocols comply with related jurisdictional privacy legislation.

## 2.4. Monitor overall data quality and resolve common data quality issues

While standardized processes may reduce the occurrence of common data quality issues (such as incomplete entries), data stewards will still need to monitor data quality so that the dataset can be used with confidence. Monitoring also ensures that any processes or tools in place to improve data quality (such as a data dictionary) are being used as intended. To assist with monitoring data quality, customized reports may be needed (e.g., to identify duplicate client records).

# SECTION 3: Introduction to the Data Lifecycle

The data lifecycle is a guiding framework used to structure and formalize data management activities. The following section describes each stage of the data lifecycle (see Figure 2) and highlights key considerations for each stage. This lifecycle framework encourages people to think about how data management activities can be formalized within an organization.

*Figure 2. Data Lifecycle Diagram*



Various data lifecycle models exist. The data lifecycle shown in Figure 2 is intended to be generally applicable to communities across Canada[[1]](#footnote-2), though exceptions may exist. Depending on the communities’ context, certain lifecycle stages may require greater attention than others.

Scenarios set within the community of Grandview illustrate the management of data in a Coordinated Access context at each phase of the data lifecycle.

## 3.1. Planning



Planning ensures that data collection efforts have a purpose and are relevant to the needs of organizations and the community. During this phase, communities should evaluate operational and strategic requirements to identify the data needs of all organizations that will participate in the data lifecycle. Data collection may be driven by the reporting requirements from funders, new programs or service delivery approaches that are being implemented or operational needs of organizations and the community.

A data management plan provides direction for carrying out the activities in the stages of the data lifecycle. The plan can be used to ensure consistency of data management practices. As data collection efforts continue, this plan can be further refined based on challenges that arise and can help ensure a consistent organizational or community mindset towards data.

|  |
| --- |
| **Planning Scenario**The community of Grandview is implementing Coordinated Access—a process to help communities connect people to appropriate services in a more streamlined way. To support this process they have hired a Coordinated Access Lead and a HIFIS Lead. They have been using HIFIS as their homelessness management information system for a few months now and want to ensure they can create a Unique Identifier List using HIFIS data to effectively serve clients. During a Coordinated Access meeting, the local Social Housing Program announced that they received funding for **30 new rent supplements with case management support.** These resources are to be allocated in the Coordinated Access Resource Inventorywith the following two prioritization criteria:1. People experiencing chronic homelessness, with additional priority to those in living rough or outside, and
2. People with the highest acuity score.

This information to help prioritize, however, is only available for a few clients. In fact, Grandview’s HIFIS Unique Identifier List shows that **assessment** information is missing for many client files. The **HIFIS Housing Status**, which indicates if a client is chronically homeless, often displays “unknown housing”. For the latter, Grandview’s Coordinated Access Lead and HIFIS Lead both know that this is due to missing HIFIS Housing History information. Also, the Leads have anecdotally heard from staff that the Unique Identifier List does not include all people experiencing chronic homelessness in Grandview.To resolve these data-related issues, the Leads develop a plan with concrete actions by determining:* What data they need to collect,
* What are the resources that are necessary, and
* Who will be responsible for data collection.
 |

**Key Considerations**

When developing a data management plan, consider the following:

* Why do I need to collect data?
* What are the data needs of my organization and the community?
* How will this data be stored and kept secure?
* What are the data collection standards or procedures that I need to be aware of, or develop?
* Do I have the necessary resources to acquire this data (both personnel, technological, and financial)?
* Who will be responsible for collecting this data?
* What skills will be necessary to acquire/transform/collect/store this data?
* Who will have access to this data, under what conditions and for how long?

## 3.2. Acquisition



The acquisition phase of the lifecycle focuses on who is responsible for collecting the data and how this data will be collected. This phase builds on the planning phase by addressing client consent to acquire the data, and additional agreements between partner organizations or employee confidentiality agreements.

Some key privacy and legal documents that support data acquisition:

* Client Consent Form, a form that discloses the purpose for which the information is being collected and used
* Community Data Sharing Agreement, an agreement with community partners that may access the data
* Confidentiality and User Agreement, a legal contract with the data users agreeing on the terms and conditions for using the data and outlines their data responsibilities.

During any data collection effort, it is important that employees understand the value of each data element being collected. Establishing data entry protocols and processes will ensure that the data is of high quality for the purposes of serving clients, further analysis, and system-level reporting.

To minimize possible data entry errors, providing training, refresher training, and job aids can remind employees what information needs to be entered and why it is important, where it needs to be entered, and in what format.

Best practices in data collection activities to ensure high quality data will be addressed in greater detail in Section 4 of this document.

|  |
| --- |
| **Acquisition Scenario**In their planning phase, Grandview has found that they have missing data related to clients’ housing history and acuity scores. Both of these elements will help ensure that the 30 new rent supplements go to clients in the most need. To determine this information, the community has decided to **conduct a joint PiT Count and Registry Week** to help ensure that their Unique Identifier List is comprehensive and complete including an up-to-date assessment and information about their current living situation.Second, they plan to have a **blitz of outreach** activities and establish clear directives to **all access points** to help fill-in any key missing information from client files. In their plan, they outline the **data points that must be collected** by certain staff roles. They communicate clearly how this client information will support referrals to appropriate community services or housing resources. For the **registry week,** Grandview maps the survey locations and train volunteers on how to record consents and conduct assessments. Trained staff are responsible for entering registry week information in HIFIS.Grandview also revises its **HIFIS** **policies and protocols** and provides training to outreach and access point workers. This helps ensure that they understand and can communicate the value of collecting this information to people experiencing homelessness. Given the clearly identified benefit in this exercise for clients, outreach and access point staff and registry week volunteers are excited about the potential impact they can have on the community. |

**Key Considerations**

For the acquisition phase, consider the following:

* Who is responsible for collecting this data?
* How should data be entered into the system and in what format (i.e., data standards)?
* Can current data systems be used to collect and store these data points?
* What consents are required?
* How to ensure that the data is good quality?
* What training will be required?
* What processes will be used to transition organizations on to a new system?

## 3.3. Processing

Data can contain errors, such as duplicates, or typographic and entry mistakes. To remove these potential errors, the data must be processed or cleaned. This will improve the quality of the data and allow for more reliable insights.

Communities undertaking data conversion, from an older data system into a new data system, should establish procedures to merge duplicate profiles and clean data. For example, an organization that has worked closely with a client with a duplicate record could be assigned the responsibility to verify the data and merge the profiles.

Having better data will allow communities to make informed decisions, resulting in better outcomes for clients and less duplication of effort for workers. It will also increase the ability to identify where gaps exist in the service delivery system.

Reviewing data from the point of entry is important to the processing phase. This ensures that all information is standardized when it enters your database and will make it easier to catch duplicates. As issues are found, create standard data entry protocols to ensure that you are only entering quality data in your system.

|  |
| --- |
| **Processing Scenario**Grandview has recently completed its registry week and outreach blitz and all information has been entered in HIFIS. The community’s access points continue to validate the information of clients, as they show up, to ensure it is up to date. Grandview’s Coordinated Access Lead and HIFIS Lead generate the HIFIS Unique Identifier List again and are eager to see if their efforts paid off. The List:- Contains an additional 219 people experiencing homelessness- Key fields supporting prioritization, such as acuity score and identification of people experiencing chronic homelessness, are now complete However, after sorting and filtering client-specific information, the Leads identify potential data quality improvements. Staff working with specific clients confirms that birthdates and episodes of homelessness seem inaccurate for certain clients. Staff follow up with clients to confirm uncertain information. This is an important step to validate the accuracy and quality of the data, which increases staff confidence in the decisions that will be informed by the List. |

**Key Considerations**

At the end of the data cleaning process, you should be able to answer the following questions:

* Does the data make sense?
* Does the data follow the appropriate rules (i.e., standard data entry protocols)?
* What kind of checks need to be performed on a regular basis to catch errors and omissions early? How to rectify them?
* Does the data bring any insight to light?
* Can you find trends in the data to help you make decisions?
* If it does not help, is that because of a data quality issue?

Section 4 of this guide describes best practices for improving data quality and can reduce the amount of data processing that is necessary.

## 3.4. Analysis and Reporting



Data analysis is the process of developing answers to questions through the examination and interpretation of data. The basic steps in the analytic process consist of identifying issues, determining the availability of suitable data, deciding on which methods are appropriate for answering the questions of interest, applying the methods and evaluating, summarizing, and communicating the results.

Data analysis also plays a key role by pointing to data quality problems in a given issue. Analysis can influence future improvements to programs, policies, or operational procedures. The type of analysis to perform depends on the characteristics of the data and the goal of the analysis. For example, quantitative data, such as numbers of clients, can be analyzed for the purpose of understanding a population’s characteristics.

Analytical results highlight the usefulness of data by shedding light on relevant issues. Reporting is used to communicate and contextualize these results and offer potential decision-making options.

|  |
| --- |
| **Analysis and Reporting Scenario**The community of Grandview has successfully improved the comprehensiveness and quality of their Unique Identifier List. Now that information on clients has been planned, acquired, and processed, the Coordinated Access Lead can confidently identify clients eligible for the rent supplements with case management support and prioritize them according to community criteria. As a result Grandview is able to allocate 24 of its 30 rent supplements.The Coordinated Access and HIFIS Leads know that all this work has other benefits: * The community has a more accurate picture of the number of people experiencing homelessness in their geographic region.
* They also know the average length of time these people spend homeless and can more easily connect with them when vacancies become available and are referred for an offer.

In order to raise the profile of homelessness issues within their community, Grandview publishes a press release—supported by homelessness data from HIFIS—to demonstrate accomplishments, while highlighting where efforts still have to be made to resolve homelessness in their community. One of their hopes is to get more landlords engaged to help house people and do some good for Grandview. |

**Key Considerations**

For the analysis and reporting of data, consider the following:

* Why do I need to analyze the data?
* What data needs analysis?
* Will the data we have collected meet our needs?
* Who will be responsible for analyzing the data?
* At what intervals will analysis be done?
* How are results going to be communicated or visualized?
* How to log and act on feedback from stakeholders?

## 3.5. Preservation and Storage



Data preservation and storage support data integrity by agreeing on a set of principles and practices when it comes to backing up your data. It involves keeping data for some period of time and/or to set data aside for future use. This includes data storage and data archiving, the practice of identifying data that is no longer active and moving it out of production systems into long-term storage systems.

There are three main types of data backup:

1. Full backup — A full backup is where every single file and folder in the system is backed up. A full backup takes longer and requires more space than other types of backups, but the process of restoring lost data from backup is much faster.
2. Incremental backup—Only the initial backup is a full one. Subsequent backups only store changes that were made since the previous backup. The process of restoring lost data from backup is longer but the backup process is much quicker.
3. Mirror backup— A mirror backup is when an exact copy is made of the source data. The advantage of mirror backup as opposed to full, incremental, or differential backups, is that you are not storing old, obsolete files. When obsolete files are deleted, they disappear from the mirror backup as well when the system backs up. The downside to mirror backup is that if files are accidentally deleted, they can be lost from the backup as well if the deletion is not discovered before the next scheduled backup.

Having your data backed-up will ensure that if a system crashes or other failures occur, your data will be secure and operations can resume as quickly as possible.

|  |
| --- |
| **Preservation and Storage Scenario**During the HIFIS implementation process, Grandview decided that backups would be done during the night on a separate server. This ensures that all data is secure and if a system fails, their data loss will be minimal. In addition, the Coordinated Access and the HIFIS Leads have decided to keep a historical record of the Unique Identifier List generated with HIFIS. To ensure the information is safe, they have established a protocol to manage historical data from the List, and designated a secured storage process to which appropriate staff have access. |

**Key Considerations**

For the preservation and storage of data, consider the following:

* How soon do I need the data back if lost?
* How fast do I need to access the data?
* How long do I need to retain data?
* How secure does it need to be?
* What regulatory requirements need to be followed?
* What policies and procedures need to be established?

##

## 3.6. Documenting with Protocols



Data management processes should be [documented](https://www.dataone.org/best-practices/document-steps-used-data-processing) for all phases of the lifecycle. Documentation supports the interpretation of results, replication of processes, and improvement of data integrity and data quality. In situations of high turnover, where a client may be assigned to a new case manager and new employees need to be trained, documentation can be used as a training tool to ensure consistency of practice. Remember to document any new processes and protocols discovered as a result of errors in management of the data throughout the phases.

**Key Considerations**

For the documentation phase, consider documenting the following:

* Data cleaning decisions
* Analysis methodologies for consistent reporting
* Questions that can be answered using the dataset
* Data check protocols
* Training protocols

Data travels through the lifecycle and is managed in different ways at each phase. How well it is managed at each of the phases influences the quality and usability of that data. The next section of the guide outlines ways to enhance and ensure high quality data.

# SECTION 4: Ensuring High Quality Data

Quality is perhaps the most important attribute of any dataset. A well-managed dataset will generate reliable and accurate results, building trust and confidence in the data as well as any insights and actions that follow from its analysis.

To support the collection and use of high quality data, written policies and protocols should emphasize the expectation of accurate, timely and complete data, all of which are necessary for sound decision-making. Written protocols should also outline the agreed-upon process for granting access to the dataset. While the dataset needs to be protected from unauthorized access, it is also important that those who need access to the data are able to do so easily. It may be necessary to refer to federal or provincial legislation to determine specific security and retention requirements for data storage.

Solutions to improve data quality are not always technical. Some suggestions include:

* Ensure relevancy;
* Set standards for consistency, such as a data dictionary and data entry guide;
* Do data quality checks;
* Invest in data literacy; and,
* Support informed use of the dataset.

## 4.1. Ensure relevancy

Datasets must have an operational and/or strategic value to justify the time and effort invested in them. Simply put, if the data has value, there is a greater chance that staff will work to maintain its quality. During training sessions about data management expectations, it is important to emphasize both the value of the data that is being collected as well as the consequences of poor quality data. For example, trainers could identify the impact of poor quality data on clients, explaining that leaving certain fields blank could result in missed opportunities for clients, such as not being able to receive a service they need.

##

## 4.2. Standards for consistency

Clarifying data management expectations through documentation of written policies and protocols will help to ensure greater consistency in how people engage with the dataset, from entry through to analysis and reporting. At a minimum, standards should be set for:

* Which data fields are mandatory or optional for different staffing roles and why;
* How each data field and drop-down option is defined;
* How to input data into the system; and,
* When or how often data entry for each field should happen.

**Example:** Develop a data dictionary: A data dictionary will help ensure a common understanding of the terminology used to describe fields or drop-down menus.

It also helps reduce the likelihood of staff making data entry decisions based on inaccurate assumptions or interpretations. The data dictionary should be easily accessible to staff and form part of new staff orientations.

**Example:** Develop a data entry guide:As different staff may be performing data entry tasks, having well-defined and standardized processes for doing so will promote consistency in the resulting data. Data entry guides can use diagrams or “process maps” to help communicate the correct elements and sequencing of a data entry process.

## 4.3. Do data quality checks

Efforts to resolve data quality issues can be more effective if common errors are identified and targeted proactively. Implementing regular data quality checks, such as generating a report every day to identify common data entry errors, can help to automate the process. It is also important to monitor whether staff are interpreting data correctly. For example, managers could review client case notes to compare consistency between staff members in how common assessment data is being interpreted.

## 4.4. Invest in data literacy

Data quality errors often originate from data entry mistakes. Providing regular training on how workers can meet data entry expectations is likely to have a significant impact in overall data quality. More training should be offered to staff with limited data management experience and to those who express a desire in becoming more confident with data. Different methods can be used for training, such as workshops, instructional handouts, webcasts, and one-on-one instruction. It is also a good idea to have informal opportunities for staff to bring their data questions to experienced team members and/or the local data steward.

## 4.5. Support informed use of the dataset

Datasets need to be accompanied by written documentation describing the intent of the dataset and other key elements, such as when the data was collected and by whom, how the data has been processed and any known issues with data integrity. Making this documentation readily available will support the informed, collective use of the dataset. This is particularly useful in environments where there is high staff turnover or in situations where analysts were not present during the data collection or processing steps.

# Conclusion

This guide illustrated some common data management issues in the practical context of a community implementing Coordinated Access and HIFIS. It also presented the concepts of a data steward and the data lifecycle, as well as provided some best practices to help communities create high quality data and manage it effectively.

Managing data takes work, but it can have far-reaching effects on clarifying collection processes, supporting accurate and valid information in HIFIS, and being able to conduct accurate analysis, all of which can lead to more effective service delivery and decision-making. Good data management practices will strengthen the housing and homelessness response systems in communities and contribute to ending homelessness in Canada.

# Resources

These resources provide further information on the topics addressed throughout this guide.

|  |
| --- |
| Value of Data in the Homelessness Sector |
| **Highlights of the National Shelter Study 2005–2016: Emergency Shelter Use in Canada:** [Infrastructure Canada—Highlights of the National Shelter Study 2005 to 2016](https://www.infrastructure.gc.ca/homelessness-sans-abri/reports-rapports/shelter-study-national-etude-refuge-2005-2016-highlights-eng.html)This publication is an example of how data collection at a national level can contribute to a greater understanding of homelessness. This publication specifically addresses emergency shelter use from a sample of over 200 shelters across the country.  |

|  |
| --- |
| Managing Consent and Privacy |
| **Privacy, Consent and Data Sharing:** <https://bfzcanada.ca/by-name-lists/>  |
| **A primer on consent in HIFIS 4 from ACRE Consulting:** <https://www.acreconsulting.ca/blog/75634-a-primer-on-consent-in-hifis-4> This blog post from ACRE Consulting offers guidance on consent specific to HIFIS 4.  |

|  |
| --- |
| Data Quality |
| **Wang, R. Y. & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of management information systems*, *12*(4), 5–33**In this academic article, the authors present a framework that captures dimensions of data quality that can be applied to industry and government.  |
| **Statistics Canada (2018). Data Quality Toolkit:** <https://www.statcan.gc.ca/eng/data-quality-toolkit> Statistics Canada provides a Data Quality Toolkit that includes data quality attributes and data quality assurance practices. This toolkit also provides quality assessment checklists for both data producers and data users.  |

|  |
| --- |
| Data Reporting |
| **Reaching Home Community Homelessness Report: Reference Guide:** <https://homelessnesslearninghub.ca/library/resources/2021-22-community-homelessness-report-reference-guide/> This guide provides definitions and guidance for completing each section of the Community Homelessness Report (CHR) template. It also identifies other resources that communities may find useful as they complete and reflect on their CHR. |
| **Unique Identifier List Report User Guide:** <https://www.homelessnesslearninghub.ca/sites/default/files/resources/Unique%20Identifier%20List%20Report.pdf>The Guide includes steps on how to access, manage and generate the report, a description of its content and calculations, as well as best practices to obtain a high quality report. |

|  |
| --- |
| Data Stewardship |
| **Data Custodianship Guidelines from the Government of British Columbia:** <https://www2.gov.bc.ca/assets/gov/data/data-management/data_custodianship_guidelines_for_the_government_of_bc.pdf> This document details the role of data custodians[[2]](#footnote-3) and outlines other roles and responsibilities relating to data governance.  |

1. . The elements of the lifecycle in this document have been adapted from the United States Geological Survey (USGS) model, found at <https://www.usgs.gov/products/data-and-tools/data-management/data-lifecycle>. [↑](#footnote-ref-2)
2. . The terms “data custodian” and “data steward” are often used interchangeably. While some definitions make distinctions between these two roles, the role of “data custodians” in this document from the Government of BC is considered to be analogous to the role of “data steward” as defined in this guide.  [↑](#footnote-ref-3)