

HUD's

HOMELESS ASSISTANCE PROGRAMS



Enhancing HMIS Data Quality

July 2005



U.S. Department of Housing and Urban Development
Office of Community Planning and Development

HUD's Homeless Assistance Programs

Enhancing HMIS Data Quality

July 2005



U.S. Department of Housing and Urban Development
Office of Community Planning and Development

Acknowledgements

This paper was prepared by the Center for Social Policy at the John W. McCormack Graduate School of Policy Studies, University of Massachusetts Boston, for the U.S Department of Housing and Urban Development's Office of Community Planning and Development under a sub-contract with the QED Group, LLC. The primary authors of this paper are Brian Sokol, Tatjana Meschede, Ph.D., and John McGah of the Center for Social Policy.

Like data quality itself, a quality document can only be the result of the combined effort of many people on various levels. Brooke Spellman, as usual, provided excellent editorial support as well as constructive contributions to the content. Michelle Hayes, Julie Hovden, and Michael Roanhouse offered valuable feedback and technical reviews. Ann Oliva was instrumental in shepherding the document across the finish line to completion.

The authors are grateful to many people working on HMIS across the country who contributed ideas for various parts of this paper, including Tom Albanese, Bryan Tosi, Sarah Donnelly, Julia Tripp, Bill Matson, Jorge Mendez, and Fred Best. Barb Ritter was interviewed extensively for the case study found in Chapter 6 and cheerfully responded to follow-up questions. Barb Ritter, Matt White, Loren Hoffman, and Susan Ma all provided reports and documents that were adapted for the Appendices. The New England Regional HMIS group acted as a pilot audience for the data standards training template that accompanies this document. We thank all of these people for their willingness to contribute their materials, time, and insights toward enhancing this paper.

Table of Contents

Introduction.....	3
Document Overview and Intended Audience	3
Chapter 1: The Importance of Data Quality	4
Definition of Data Quality	4
Importance of Data Quality for HMIS Goals	4
Types of HMIS Data.....	5
System Generated Information	6
Advantages and Limitations of HMIS Data.....	6
Chapter 2: Data Quality Issues and Standards.....	7
Timeliness of Data	7
Data Completeness.....	8
Data Accuracy (Data Validity)	12
Data Consistency	13
Data Quality Plans	14
Chapter 3: Program-Level Staff.....	16
Front-line staff	16
Data Entry Staff	21
Agency and Program Directors.....	24
Chapter 4: Implementation and Continuum of Care Staff	27
Mechanisms Prior to Entering Data.....	27
Data Quality Plans	27
Consistency Among Agencies	27
Validating and Cleaning Data.....	28
Releasing Data	29
Chapter 5: HMIS Software Solutions	30
Screen Design	30
Field Labels.....	30
Response Options.....	31
Validation.....	32
Flags for Missing or Stale Data	33
Non-Typing Technologies	34
Reports and Logs	34
Chapter 6: Putting the Pieces Together: A Case Study of Michigan State HMIS.....	37
Overview.....	37
Philosophy.....	38
Data Quality Standards	39
User and Agency Level.....	39
Continuum of Care level.....	41
Statewide Implementation Level	41
Challenges.....	42
Conclusion	44
Chapter 7: Releasing Data	45
Types of Program.....	46
Data Fields and Sub-Groups	46

Appendix 1: Training Template.....	49
Appendix 2: Universal Data Elements Data Entry Form.....	50
Appendix 3: Data Quality Plan Work Sheet.....	54
Appendix 4: Sample Data Quality Log.....	57
Appendix 5: Program Level Data Process and Feedback Loop	58
Appendix 6: Director’s Memo	59
Appendix 7: Possible Validations of HMIS Universal Data Elements.....	60
Appendix 8: Sample Data Quality Reports.....	62
Aggregate Implementation-Level Report	62
Single Score Per Program Report	63
Single Program Report Including Additional Guidelines.....	64
Aggregate Reports by Program for HMIS Project Manager	66
Single Program Data Quality Report Sent to Program Directors	67
Appendix 9: Documents from Michigan State HMIS	69
Excerpts from Policies and Procedures.....	69
Quality Assurance Policy and Procedure (2004)	71
Appendix 10: Sample Statement of Limitations, Coverage Rates, and Parameters	75

Introduction

Collecting quality information in any type of social research is a challenging task; it is especially challenging when collecting data on people experiencing homelessness. Faced with daily tasks such as looking for food, seeking shelter and income, or caring for their children, providing information to be entered into a homeless management information system (HMIS) is not one of their chief interests. Distrust of the social service system and mental health issues may further prevent many from disclosing personal information. In addition, homeless service staff members often tend to focus on direct services and not on quality data collection. However, to fully understand the nature and extent of homelessness, associated service needs and service utilization patterns, it is critical to collect the most accurate and representative information on individuals and families who experience homelessness. Finding solutions to homelessness requires quality system-wide, longitudinal data. This paper will describe program and system level strategies to improve data quality.

Document Overview and Intended Audience

This document is divided into seven chapters. Chapter 1 focuses on the importance of data quality for achieving HMIS goals. It also discusses the types of data captured by an HMIS and the advantages and limitations of HMIS data compared to other types of information. Chapter 2 outlines issues critical to developing data collection and entry standards. This chapter contains ideas that are relevant to all HMIS stakeholders.

Data quality is a joint effort. Chapters 3 through 5 discuss the various parties that are instrumental for ensuring data quality. Chapter 3 focuses on program level staff, including front-line staff people who collect the data, data entry staff, and program directors who are responsible for fostering data quality efforts and for monitoring quality in the program. Chapter 4 is aimed at HMIS project staff, including HMIS project managers, system administrators, technical assistants, and CoC conveners who manage and/or use the system level data. Chapter 5 discusses the various ways in which the software itself can contribute to data quality. This chapter targets software developers, system administrators, trainers, technical assistants and users who may be able to identify new ideas, take advantage of advanced features, ask for enhancements, and be aware of potential strengths and pitfalls of particular solutions. Those in the process of selecting software should also find this chapter useful.

Chapter 6 is a case study of the State of Michigan HMIS, which shows how one implementation has been able to put all the pieces together to enhance data quality.

Chapter 7 briefly discusses issues related to releasing data.

The paper is peppered with “Tip Boxes” highlighting useful ideas. As well, the document makes repeated reference to the importance of training HMIS stakeholders in various aspects of data quality. To facilitate this effort, Appendix 1 is a PowerPoint presentation that includes many of the lessons contained in this paper. The training materials can also be downloaded in electronic form, adapted and used as part of a basic HMIS training curriculum. Other appendices include valuable sample resources, such as a sample data quality “cheat sheet” and data collection instrument, a data quality plane worksheets, a list of validation checks for particular fields, and sample data quality reports.

Chapter 1: The Importance of Data Quality

Definition of Data Quality

HMIS data quality refers to the extent that data recorded in an HMIS accurately reflects the same information in the real world. A perfect overlap between data and reality would result in a hypothetical data quality rating of 100 percent, while a data quality rating of 0 percent would indicate that there is no match between the information entered into an HMIS and the same information in the real world. No data collection system has a quality rating of 100%.¹ However, to meet the HMIS goal of presenting accurate and consistent information on homelessness, it is critical that an HMIS have the best possible representation of reality as it relates to homeless people and the programs that serve them. Specifically, it should be our goal to record the most accurate, consistent and timely information in order to draw reasonable conclusions about the extent of homelessness and the impact of homeless services.

Importance of Data Quality for HMIS Goals

Data quality is greatly improved when the goals of data collection are clear. The goals of HMIS on a national level were stated by Congress:

There has never been an overall review or **comprehensive analysis** on the **extent of homelessness** or **how to address** it. The Committee believes that it is essential to develop an unduplicated count of homeless people, and an **analysis of their patterns of use** of assistance ...including how they enter and exit the homeless assistance system and the **effectiveness of assistance**.²

Thus, the 2001 Congressional directive targets information to understand:

- The extent of homelessness,
- The nature of homelessness (implied in “comprehensive analysis” and necessary to know “how to address”),
- Homeless service use patterns, and
- The effectiveness of the homeless service system

These goals are not only important on the federal level but also critical for understanding homelessness and program planning at local levels.

Extent of Homelessness

The number of homeless people has been at the center of debate for as long as homelessness has been acknowledged as a social problem. Due to inconsistent or no data collection, different estimation methods result in largely diverse numbers. One goal of HMIS is to estimate the number of homeless people that closely represents reality. By collecting personal information on

¹ Ken Orr, Data Quality and Systems Theory, Communication of the ACM, February 1998/Vol. 41, No. 2, pp 66-71

² Senate Report 106-410.

all clients served, HUD hopes to generate an estimate of the unduplicated count of homeless people that access services nationally. Achievement of this goal depends on high quality personal identifying data, such as Social Security Number, names, gender and date of birth, which are used to create unduplicated counts.

Nature of Homelessness

Additional HMIS data elements focus on the characteristics of those engaged in homeless services. Analyzing this information on a larger level will improve our understanding of the people experiencing homelessness, the issues they face, and their service needs. High quality data on gender, date of birth, race, ethnicity, veteran's status and disability, and household composition are needed for this goal.

Pattern of Homeless Service Utilization

People who are homeless often use more than one of the programs that are available to help them access housing, resolve their crisis, support them, and link them with other services. Accurate program entry and exit dates and information on residence prior to program entry are critical in determining service use patterns that assess average length of stay and movement among different homeless programs. The collection of accurate identifying information at each program is also necessary in order to identify the extent to which clients appear in multiple programs, how clients move through the system, and to detect cycles of homelessness

Effectiveness of the Homeless Service System

Assessing the effectiveness of current the homeless service system is critical to finding successful solutions to ending homelessness. For that reason, information at program exit, such as destination and income, are important to learn if and how the system has helped to resolve clients' housing crisis and to improve their overall stability. Data on returning clients also contribute to this goal. Comparing program entry data with program exit data at the aggregate level will also provide a picture of homeless program impacts on the clients they serve.

Types of HMIS Data

HMIS systems collect three basic types of data: self-report information by homeless service users, observational data by service staff or clinicians, and system-generated information.

Self-Report

Most of the data entered and stored in HMIS systems are based on self-reported information by homeless clients. This information includes identifying information (e.g., name, birth date), demographic information (e.g., race/ethnicity, educational attainment), and assessment information (e.g., past residences, needs, and services received). The accuracy of self-report data may be limited in terms of collecting valid information for some data fields, such as substance abuse or mental health status (see Chapter 3 for further discussion of this issue).

Observational/Clinical

Some of the self-report data can be supplemented or validated by observations of homeless service staff or clinicians. For example, if an individual does not share any information regarding substance abuse or mental health issues, clinical diagnoses can be entered into the corresponding data fields. The HMIS Data and Technical Standards Final Notice specifies which required fields may be collected through observation. Among the Universal elements, Disabling Condition may be collected through observation, but other fields, such as race, may not.³ In general, data collection should not be surreptitious, and all data collection should be done within the confines of the privacy standards and local policy.

System Generated Information

HMIS systems can generate a number of different types of information. For example, based on accurate program entry and exit information, HMISs can generate information on the length of time an individual or family has been served by the homeless service system. In addition, service use across different types of programs can be generated. Lastly, data on where clients stayed the night before program entry can be validated, if the client used another of the programs contributing data to the HMIS.

Advantages and Limitations of HMIS Data

Administrative data sets, such as HMIS data, have the advantage of collecting information on all clients served, as compared to surveys which sample a subgroup of individuals for data collection. As such, with close attention to data quality, HMIS information has great potential to capture very valuable information on homeless clients. However, as with any administrative data set, there are also inherent limitations to HMIS generated data sets. For example, HMIS data cannot describe all of the circumstance by which homeless people became homeless, nor can it present information on people who are not accessing services, or evaluate certain program activities. It may not be practical to collect specific information on all clients in order to generate representative data to answer a specific research question. Other data collection methods, such as point-in-time surveys or ethnographic studies are better suited to answer these questions.

Information on homelessness can be enhanced by employing other data collection methods in addition to HMIS. Further, linking HMIS data sets with other administrative data sets, for example TANF, would provide the opportunity to assess long-term outcomes for homeless people who have moved out of the homeless service system. In sum, to gain a more comprehensive understanding about homelessness, surveys, qualitative data collection, and linkage with other administrative databases supply important sources of information that better help communities and Congress to understand homelessness and lead to solutions to ending it.

³ Department of Housing and Urban Development, Homeless Management Information Systems Data and Technical Standards, Final Notice, 69 Fed. Reg. 45889-45934 (July 30, 2004). (Hereafter *HMIS Final Notice*). Secs. 2.4.2; 2.7.

Chapter 2: Data Quality Issues and Standards

One of the most effective ways to collect quality data is to develop data collection and data entry standards that are implemented by all programs entering data into the HMIS. These standards will ensure that data are entered in a timely fashion and consistently across different programs. This chapter presents guidelines for timely, complete, accurate and consistent data entry and monitoring approaches. Information on who should be responsible for adhering to these guidelines will be outlined in the next chapter.

Timeliness of Data

To be most useful for reporting, an HMIS should include the most current information on the clients served by participating homeless programs. To ensure the most up to date data, information should be entered as soon as it is collected. This is not a problem when data are entered directly into a database and not collected on paper. However, entering data directly into a computer may raise other concerns. (See Chapter 3 for a fuller discussion of the pros and cons entering information on paper versus directly into the HMIS). CoCs should specify concrete timelines for when data entry of information collected on paper should be completed. For example, the Columbus Community Shelter Board specifies in their HMIS Quality Assurance Plan⁴ that intake data need to be added “within 24 hours of the intake process” The State of Kentucky has a similar requirement. Other CoCs have varying requirements that specify that data entry must be within two days (Houston), three days (Fort Pierce), or seven days (Chicago) of intake or client encounter.

Information that tends to change periodically also needs to be regularly verified and/or updated., such as information on income sources and amounts. Recommendations in HMIS protocols on when to update these data vary. Columbus’ HMIS Quality Assurance Plan specifies that information other than intake data needs to be updated monthly (see below).⁵ Other HMIS protocols state that data on people served in the previous month need to be updated by a specified date in the following month, ranging from by the 4th day to the 15th day.⁶ Chicago’s agency participation agreement specifies that data on a client should be entered within seven days of client interaction, which covers initial intake and subsequent service encounters. Some software solutions facilitate this by flagging clients whose data has not been updated for a certain length of time (see Chapter 5).

Exceptions to the timeliness principle can be made for domestic violence providers, which may wait until clients leave the shelter before entering data into the HMIS or sending data for analysis. In these cases, the standard may be that the data are entered within a specified period after the client leaves, rather than after intake, as long as client safety is ensured.

⁴ Community Shelter Board, “Client Tracking and QA Standards”, Columbus, Ohio. http://www.csb.org/Agency_Resources/agency_resources.htm.

⁵ Community Shelter Board.

⁶ Community Shelter Board; Michigan Coalition Against Homelessness: <http://www.mihomeless.org/>

Excerpt from Columbus HMIS Quality Assurance: I. Client Tracking & QA Standards

D. Reporting Submission Deadlines:

1. Intake data should be entered into the HMIS **within 24 hours of the intake process.**
2. Shelters only: Clients who stayed in shelter during the previous 24-hour period must be entered into [the] Bed List daily by 9:00am.
3. Complete and accurate data for the month must be entered into HMIS by the **fourth working day of the month** following the reporting period. For example, data for the month of April must be entered into HMIS by the fourth working day of May.

Data Completeness

To release meaningful information from the HMIS, data need to be as complete as possible, i.e. they should contain all required information on all people served in a certain type of program (i.e. emergency shelter) during a specified time period. On the macro level, the goal of achieving adequate HMIS coverage and participation by all local programs is essentially about ensuring that the records are representative of all the clients served by these programs. When individual records or whole programs are missing, it is important to consider whether the characteristics of those served by the missing program are significantly different than those that are included. If a client record is missing, then aggregate reports may not accurately reflect the clients served by the program. Similarly, if an entire program is missing, data from the HMIS may not accurately reflect the homeless population in the community.

Missing Client Records

Even with all programs participating, it is possible that not every client served by the program is actually being entered.⁷ Missing client records from participating programs is particularly problematic since, unlike missing programs, the extent of those missing is difficult to quantify, and such gaps will not be factored into the extrapolations used to generate the overall homelessness count. That is, while it is possible to know what percentage of beds are represented by participating and non-participating programs and adjust estimated counts accordingly, it is much more difficult to say within a particular program what percentage of clients are not being entered. In addition, like with missing programs, missing clients within a program might have characteristics that skew the data findings. For example, those who stay only one or two nights might be more likely to not be entered. If this is the case, aggregate length of stay information can be severely skewed toward longer stays.

Tip: Setting data entry deadlines of one or two days after intake ensures timeliness and avoids rushed, sloppy, burdensome data entry at the end of a reporting period.

One strategy to address the issue of missing client records is to compare paper records (i.e. manual nightly shelter check-in lists) with the information entered into the HMIS, which should reveal any missing client records.

⁷ The coverage calculations used by the Annual Homeless Assessment Report are based on the assumption that 100% of clients served by participating programs are entered into the HMIS.

Incomplete Client Records

The second type of incompleteness in a dataset is missing fields within particular client records. Standards should be set to ensure that all required fields are consistently answered. For example, the Columbus Community Shelter Board clearly defines the level of accepted missing information. In their Client Tracking and QA Standards (see below), they state that the level of missing or unknown information should not exceed 5% for any given data field in any given month. Clearly stating an expectation on data completeness will lead to increased data quality and more valid, relevant and policy impacting reports. Where possible, if clients do not know or refuse to answer a particular question, this should be stored as an answer in the database, rather than leaving the field empty.

Excerpt from Columbus HMIS Quality Assurance: I. Client Tracking & QA Standards

E. Data Accuracy:

1. All clients have unique ID numbers (Social Security Number or system-generated ID).
2. Missing/unknown data in HMIS is **less than 5% per month in required variable fields**. For example, if the data for the variable veteran is unknown for less than 5% of clients during the month, the data is accurate. If unknown is greater than or equal to 5%, the data is inaccurate. The **only** data variable exception to accuracy, with respect to 'Unknown' is the variable Destination.
3. No data incompatible with program in HMIS. For example, a family cannot be entered at a single men's shelter or a women's shelter.
4. Data in HMIS must accurately reflect client data recorded in the agency's client file and known information about the client and services provided to the client. For example, 'Exit Date' on the Columbus Worksheet should be the date the client physically exited the shelter.

There are two main approaches to ensuring that all required fields are completed consistently: software validation and data quality reporting:

1. With software validation, records are not saved unless all required fields are entered. This approach is effective at capturing something for every field, but may also lead to staff entering inaccurate information just so they can save the data. If this approach is used, it is vital to include "don't know" and "refused" options for required fields. A good compromise is for the software to generate a warning before saving fields with incomplete (or invalid) data.
2. Data quality reporting occurs after the fact where an agency or system administrator produces reports of missing fields, and feeds that information back to the agency and data entry staff. Quality reports can be aggregate, producing a percentage of completeness for each field on an agency, program, or user level (e.g., User A completed the "race" field for 85% of new records.). These reports can be useful for assessing overall compliance with the standards, identifying training issues and/or software design issues, and addressing programs or users that are not meeting the standards. Quality reports can also be done on the client level. In this case, actual client lists are generated that highlight which data are missing for which clients. These reports are more useful when staff is able to go back and actually fill in the missing records.

(More information on monitoring and enforcing completeness through software features and data quality reporting can be found in Chapter 5. Sample data quality reports can be found in Appendix 8.)

In addition, as with missing records, the distribution of missing responses may not be assumed to match the distribution of captured responses. This is particularly true for “Yes/No” type questions. For example, if the question asks whether the client is a veteran, data entry staff may be consistently checking the “Yes” box for veterans, but often leaving the field blank if the client is not a veteran. The result of this practice would be a very low response rate for the question, and skew toward a high percentage of veterans showing up in the data set, if missing data are eliminated for the percentage calculations. Alternatively, data entry staff at a veteran shelter may also be ignoring the “Veteran” question, since every client they deal with is a veteran and the question seems superfluous.

Tip: Data for active clients should be reviewed and updated at least monthly.

Client sensitivities, in addition to data entry shortcomings, may also lead to an uneven distribution of missing responses. For example, clients who actually have disabilities may be more likely to refuse to answer questions about whether they have disabilities. Similarly, it is much more feasible to conduct exit interviews and collect destination information for clients who completed a program and had successful outcomes, then for those who returned to the streets and simply did not show up one day.

Records need to be regularly checked for their completeness. Most likely, basic client characteristics are entered during intake. Missing fields affects the ability to generate statistics about the specific field; therefore, procedures need to be in place on when to add other information to the client record, such as income and health status information. As pointed out earlier, the later information also needs to be updated regularly. Depending on the type of program, these updates should be conducted at least monthly.

While most HMIS implementations collect valid intake data, including date of program entry, program exit information is often incomplete or missing all together. However, this information is critical in order to assess service utilization patterns and outcomes associated with service use. Without program exit information, service use records are incomplete. Procedures need to be in place to ensure that program exit information is collected and entered into the HMIS. Program exit information is also necessary for calculating both length of stay and determining who is being served during a particular period.

Tip: The highest standards should be applied toward achieving data completeness for the identifying fields: name, social security number, date of birth, and gender.

Domestic violence providers who may withhold identifying information should apply equally high standards to ensure that the coded information they send is complete and accurate.

Incomplete Identifying Information

Incomplete client identifying information – specifically, name, Social Security Number, date of birth, gender, and household identifiers – will impede the Continuum’s ability to determine unique clients, hinder the client matching process, and throw

off the unduplicated count of clients and households. If insufficient data is provided, it is impossible to generate unique IDs and to verify whether two records represent the same client; thus, the count could appear higher than it is in reality. It could also be lower than it should be, if for example, there are two clients with the same name, but no Social Security Number is recorded for one of the clients. The HMIS or data analyst might assume they are the same client. However, a Social Security Number could have proven that they were different clients.

Both higher and lower counts can have untoward consequences. If counts are too low, the scope of the problem is understated, and the amount of resources directed to homeless programs could be lowered. If the count is too high, the successes of the service system in reducing homelessness are minimized. This, too, can affect resource allocation. For this reason, it is best to concentrate on generating the most accurate number possible, which necessitates collecting quality identifying information.

Unfortunately, identifying information is most closely linked to concerns about client privacy and confidentiality, making collection of these data that much more difficult. Even though the HMIS software may allow for anonymous data entry, this practice is directly linked to poorer data quality. As such, this practice should be avoided to the extent possible, since it automatically throws off the unduplicated count. There are other methods can be employed to protect client privacy and safety without compromising the quality of the data.

The highest standards should be applied toward achieving data completeness for all the fields used for unduplicating clients. The CoC might set a 95% standard of completeness for identifying fields, while, at least initially, somewhat lower standards of completeness might be expected for the other fields. Community and program level data quality initiatives might target these fields first. These fields can also be emphasized by placing them prominently on the intake screen, validating for them in the software, and making special efforts to fill in this information retroactively if the fields are missing.

Domestic violence may use “a proxy, coded, encrypted or hashed unique identifier—in lieu of name and SSN.”⁸ In this scenario, these providers would be excluded from the completeness goals related to identifying information. However, high standards should be applied to ensure the accuracy and completeness of the alternative identifiers used to unduplicate records.

Homeless families also need to share a unique Household Identifier in order to link all of their members for analyses. If this information is missing, it is impossible to get accurate counts of families served, data on family composition will be invalid, and each family member may be incorrectly counted as a single individual served. For example, suppose a family of four entered a shelter, but the household identifier was not generated properly. Depending on how the analysis is done, they might be counted as four families, zero families and four unaccompanied individuals, or the records might be discarded. Although the Household Identifier itself is usually system generated, users must enter clients in a particular way in order to ensure that the clients are related properly.

⁸ HMIS Data and Technical Standards Final Notice; Clarification and Additional Guidance on Special Provisions for Domestic Violence Provider Shelters 69 Fed Reg. 4848-N-03 (October 15,2004).

Data Accuracy (Data Validity)

Information entered into the HMIS needs to be valid, i.e. it needs to accurately represent information on the people that enter any of the homeless service programs contributing data to the HMIS. Inaccurate data may be intentional or unintentional. In general, false or inaccurate information is worse than incomplete information, since with the latter, it is at least possible to acknowledge the gap. Thus, it should be emphasized to clients and staff that it is better to enter nothing (or preferably “don’t know” or “refused”) than to enter inaccurate information.

Intentionally False Information

There are many reasons why clients may provide false information. These include not wanting to be tracked, general privacy issues, vanity, embarrassment, paranoia, a desire to qualify for a particular service, fear of being turned away, or simply not caring enough. In addition, caseworkers may also opt to enter untrue information to help clients, because of time limitations, or lack of full knowledge.

Educating users about the benefits of the HMIS, ensuring there are privacy and security policies in place to protect data, creating operational uses of the data that directly improves services for clients, and developing trust between clients and Front-line staff can often mitigate the amount of false information provided. Also awareness of the options of saying “don’t know” or refusing to answer is important, since these answers are generally preferable to false answers.

In addition to trainings on the importance of entering correct data, false information can be addressed through thorough data entry checks by third parties. The extent and types of false information in reports can be addressed after the fact by sharing results with stakeholders including data entry staff and consumers. Focus groups of consumers viewing the data may be able to identify areas where clients are inclined to be misleading.

Since eliciting true information is the responsibility of the front-line staff, Chapter 3 discusses these issues further.

Unintentional Errors

There are a number of unintentional errors that can occur during intake and data entry. These include:

- Accidentally selecting wrong response from drop-down;
- Misspelling (based on not knowing the proper spelling);
- Transposition of characters, or missed keys (accidental typographical errors);
- Swapped fields (e.g., first name in last name field, or intake date in exit date field);
- Use of nicknames instead of real names;
- Inaccuracies based on misunderstanding the question;
- Hearing the wrong information; and
- Transcription errors, including the inability to read handwriting.

Tip: The likelihood of data entry error increases when data are collected and entered by different staff.

Providing clients with access to review and correct the personal information that has been entered in the HMIS can improve data accuracy. This is also a client's right, as published in the *HMIS Final Notice*.⁹ Clear procedures need to be set up to allow for access to HMIS data, as well as a shared understanding of staff on how to handle such requests and use them as an opportunity to verify data accuracy. The likelihood of data entry error increases when data are collected and entered by different staff. Data entry staff people who have not personally collected the information from clients have a reduced ability to recognize data collection errors from the data collected on paper. Similarly, if significant time elapses prior to data entry, staff may not recall the notes and unintentionally enter incomplete or inaccurate data. As such, it is advisable to either have the same staff collect paper records and enter the information within a very short period of time or enter data right into the HMIS.

These issues and ideas for mitigating them are discussed further in Chapter 3. Chapter 3 also includes a discussion on advantages and disadvantages of data collection on paper as compared with directly entering data into the computer.

Data Consistency

Consistency of data collection and data entry refers to a shared understanding of what data need to be collected and in which way. Different interpretations of how questions for data collection should be asked or what answers to questions mean lead to aggregate information that cannot be correctly interpreted and presented.

For example, the question on residence prior to shelter entry has been interpreted by clients and staff in many different ways. Some thought this question referred to where individuals or families resided before losing their residence, others thought it referred the place where clients spent the night before accessing a shelter, and some may have given information on where they stayed in between. Given this range of different meanings, the information collected in this data field could not be correctly interpreted. As such, the *HMIS Final Notice* clarified the meaning of residence as referring to the night before accessing shelter, and information on where clients lived before losing their residence is collected in separated data fields.¹⁰

Tip: HMIS user trainings should clearly address which fields are required and the meaning of all required data fields.

To avoid misunderstanding of the interpretation of certain data fields, data collection and data entry staff in all agencies need to attend trainings that clearly address the meaning of all required data fields included in the HMIS. The *HMIS Final Notice* provides the basis for such trainings for the required HMIS data elements. If CoCs add other locally required fields in the HMIS, the data elements need to be clearly defined to ensure consistent data collection across different staff and across different agencies. (See Appendix 1 for the “HMIS Data Quality Training Template” and Appendix 2 for a sample form for the universal data elements, with sample questions, valid responses, notes and tips.)

⁹ . *HMIS Final Notice* Sec 4.2.5.

¹⁰ *HMIS Final Notice* Sec. 2.8.

A standardized list of all required data elements needs to be prepared and communicated to all staff involved in data collection and data entry at all agencies. At minimum, the list should include the Universal data elements listed in the *HMIS Final Notice*, but may also include the program specific data elements and/or other locally defined fields. Such a list can be used as a “cheat sheet” or checklist by staff to help ensure that they collect complete client data for each client. In addition, reminders in the HMIS software can help to make sure that all required information is entered.

Data Quality Plans

A data quality plan translates the above standards into concrete benchmarks.¹¹ For example, the plan takes the standard of “timeliness” and defines a concrete set of rules, such as “all data must be entered within two days of intake” and “all open cases are reviewed and updated by the fifth of every month.” Rules for completeness might set a benchmark that 100% of client records should be entered and each required field should be completed for 95% of records. Accuracy can be defined in terms of the percentage of fields incorrectly entered. Consistency is somewhat harder to benchmark, but can be checked by periodically discussing data elements with staff at HMIS user groups. A worksheet for developing a data quality plan can be found in Appendix 3.

In order for the benchmarks to be useful, the plan should also build in monitoring procedures, describing how the benchmarks are verified. For example, the timeliness benchmark might be verified by running a report comparing date of data entry to the program intake date entered. (The program entry date can also be compared to paper files to ensure that the dates were entered honestly.) Alternatively, the same benchmark could be verified by conducting random spot checks of programs to see if there is a backlog of clients to enter. The monitoring procedure should establish exactly who should do the monitoring. This could be the job of the system administrator. An alternative is for the agency-level administrator to do it and send the results to the HMIS project manager. The plan should also specify how often each of the various monitoring procedures should be conducted. Appropriate benchmarks and monitoring procedures may differ for different types of programs. For example, high-volume shelters may require somewhat lower standards but more frequent monitoring than low-volume shelters.

The third component of the data quality plan is a clear incentive for achieving the benchmarks. On the system-level, these incentives could be tied to funding (either positively or negatively). On the program or individual level, achievement of benchmarks could be tied to job performance reviews. These strong measures can be accompanied by additional measures such as publicly honoring individuals or programs that achieved the benchmarks. Compliance with the procedures in data quality plans should be

Components of a Data Quality Plan

- Concrete benchmarks for timeliness, completeness, and accuracy (consistency if possible)
- Monitoring procedures specifying, how and when monitoring occurs and who is responsible.
- Clear incentives for compliance
- Contractual buy-in and agreement mechanism

¹¹ See “Community Spotlight: HMIS Quality Assurance Plan for Community Shelter Board in Columbus, Ohio”, *HMIS.Info Newsletter* (January 2005). http://www.hmis.info/news_docs/January%202005%20Newsletter.pdf

considered part of the HMIS participation agreements, along with compliance with privacy standards and other rules.

Tip: Quality plans can be created for any level: implementations, programs, and individual staff members.

Data quality plans can be drafted for the overall HMIS project as well as on the program and individual levels. The plans can also include allowances for the circumstances of particular programs. Most notably, different rules may apply to domestic violence providers than other types of providers. Even where system-wide quality plans exist, programs can be encouraged to draw up their own internal plans that exceed the minimal

threshold. Data quality plans can also be individualized. For example, intake and data entry staff might have the benchmarks built into their job descriptions and monitored as part of standard supervision. Slightly different benchmarks can be created for novices and experts.

Issues specific to the various types of program level and implementation level staff in achieving data quality are discussed in the forthcoming chapters.

Chapter 3: Program-Level Staff

Achieving data quality is an ongoing, team effort. There are five key contributors to this goal: Front-line data collection staff, data entry staff, program executive staff, CoC-level staff, and the software itself. The next three chapters will look at the role of each of these partners in achieving data quality in the first instance and validating data once it is entered.

It is essential that all staff throughout the agency have a shared understanding of the need and process for achieving data quality. This chapter looks at the roles and issues different staff within a program should consider in regards to data quality.

Front-line staff

The foundation of data quality lies with front-line staff. Front-line staff members are the first people to collect information from a person receiving homeless services; they also ascertain where to put it, and then record it. In addition to the intake stage, front-line staff may also gather data throughout the client's participation in the program, at exit, and at particular follow-up points. These individuals may also enter the data (see section on data entry staff) but this section focuses on aspects of sound process and understanding to increase the accuracy of the data.

Shared Understanding of Purpose / Process

All front-line staff as well as other key stakeholders in the collection, analysis, and dissemination of data should have a shared understanding of the purpose of the data collection (e.g., to document effects of policy change, to support claims to funders, to better shape services to client needs, to understand trends across the region), and the overall process to meet those goals (e.g., front-line staff collect and record information, data entry staff enter into computer, data is "cleaned" for accuracy, data is analyzed, and reports are generated for distribution). The process could be delineated in the HMIS Policies and Procedures for the HMIS project. At minimum, all staff should have access to a written memo outlining the data collection process and explaining the importance of accurate data and maintaining data quality. Documenting the process also conveys a sense of the importance of assuring sound data.

Establishing a Rapport with Consumers

Data in most HMISs are self-reported by people seeking homeless services. Often people in the vulnerable position of being homeless may give incorrect information intentionally or unintentionally for a host of reasons. Inaccurate information can be minimized by establishing a rapport with the consumer.

Tip: Good eye contact, a warm tone, and conveying an appreciation that the information requested can be very personal, sensitive, and private all contribute to trust.

In an emergency shelter, intake is not the ideal time to ask for personal information. He or she may be disoriented or nervous. Ideally an intake worker collects only the minimally required data needed to assign a bed or a service. Once the person is settled in the shelter and has his or her bearings, the front-line staff may have more success in building a

relationship. It often helps to explain fully why questions are being asked and what will be done with the information. If this information is shared clearly and respectfully consumers are more likely to share accurate information. The rapport developed, even in a short time, can make all the difference. In non-emergency settings, front-line staff may have more time to let a person get settled before asking assessment/intake questions. In both cases, explaining the confidentiality procedures and security practices of the agency is essential and often required by law or local policy.

The manner in which questions are asked is critical in establishing a good rapport and getting accurate information. Experience suggests that basic respect and courtesy makes a big difference. Someone seeking homeless services is likely to be vulnerable, perhaps scared and feeling disconnected. Good eye contact, a warm tone, and conveying an appreciation that the information requested can be very personal, sensitive, and private all contribute to trust. Assuring clients that this information is intended to better serve them is also important.

Consumers often are not aware of the critical connection between funding and services. Communicating why the client's information is being collected, how it will be used, and how it helps the agency secure and sustain funding for the program may also be a valuable way to build understanding and support from the client. It is advisable for all staff to agree on a minimal level of information that all clients should receive. The CoC or agency may want to write out talking points and/or train users on how to consistently explain the HMIS and data collection.

Gathering True Information

It is the responsibility of front-line staff to collect and record true information from clients. Clients may be suspicious or paranoid of having their personal information entered into a computerized data system and may supply false information. Clients may supply false identifying information if they do not want to be tracked. They may also supply false information about age, prior living status, disability, pregnancy, or income, for privacy reasons, or out of embarrassment or vanity. Clients should be informed about the privacy and security procedures, and the allowable uses of the data. Explaining the goals of HMIS and how the data collection system can support individuals' access to services may also help overcome this barrier to accurate information.

Though clients should be encouraged to answer the questions, they should also be informed that no answer is preferable to a false answer.

Clients may also want to give the most advantageous answer, and believe that providing a false answer (e.g., stating a lower income) will entitle them to additional benefits, or save them from

Reasons for providing false information:

- Privacy (not wanting to be tracked)
- Embarrassment / modesty
- Paranoia
- Desire to qualify for service
- Fear of being turned away
- Not caring

Reasons for providing true information:

- Improved direct services
- Benefit eligibility and info validated
- Want to tell their story
- A relationship has been created
- Understand privacy/security procedures
- See benefits of HMIS for homelessness
- Given the option to not answer

an undesirable outcome, such as being turned away from shelter. If this seems to be occurring, staff should emphasize the goals of HMIS as well as the reasons that data are collected. When possible, staff should note any third-party documentation that has been provided for verification purposes. Finally, some clients may just not care and provide whatever answer occurs to them. It may or may not be obvious to the front-line staff when this is occurring. But, a trained interviewer is often able to tell. In any case, staff should be discouraged from entering false information.

In addition to false information provided by clients, staff may try to help their clients better access services may by recording incorrect information. Staff may choose the first answer from the pick list, if time is short. Or they may find it easier to estimate a birth date or automatically record “No,” rather than ask a sensitive question. They may also enter information that they believe will best serve the client. Finally, when staff does not know an answer, they may out of the best intentions decide to use a “placeholder” (e.g., use of “Boy” or “Girl” in a child’s first name field, when name is unknown). These practices should be avoided.

Avoiding Inconsistencies and Unintentional Errors

In addition to false information, front-line staff also should also be on guard against unintentional errors or inconsistencies. Several types of unintentional errors can occur during the intake process (as opposed to the data entry process).

The first type of unintentional error occurs when the client misunderstands the question. A common example of this is misunderstanding what is meant by the “Prior Residence” question. A client might wish to give the residence where they lived for years prior to the night before coming to the shelter as opposed to the place they stayed for one night prior to shelter entry. The actual meaning of “disability” is also easily misunderstood. Inconsistent interpretation is also a problem with these fields. Two people with the same condition might give divergent answers regarding whether they have a disability. It is up to the front-line staff to query further to determine whether which answer is most appropriate.

Tip: Always confirm the spelling of client name. “Smith” could sometimes be “Smythe.”

Language barriers can also contribute to misunderstanding the question. If many clients speak only Spanish, for example, it is helpful to have a copy of the questions and answers in Spanish available, so clients can read along. Staff members may also sometimes hear the wrong answer, especially when working with clients with strong accents or language barriers. But, this could also be a problem even without those constraints. It is quite easy to hear “No” when someone says “Don’t Know.” The intake space should be quiet and private to ensure that staff can hear clearly and follow up on sensitive questions to make sure they understand the response.

Use of nicknames and aliases is another place where misunderstanding and inconsistency causes problems. Clients who are asked “What is your name” are more likely to provide the name by which they are called than their legal name. Consistency problems occur when the client gives their legal name in one interview and their nickname in a second interview. Misspellings of names and other fields are common but easy to guard against by following a simple rule of

always confirming the spelling of a client names. Even a common name like “Smith” could sometimes be “Smythe.” Of course, spelling of names could be misheard. Circling or highlighting an unusual spelling will ensure that the data entry staff notices it.

Recording Information: The Paper vs. Computer Dilemma

There are two ways to record information during an interview: writing the information on **paper** to be entered later into a computer *or* entering directly into a **computer**. There are advantages and disadvantages to both.

Recording information on paper can lend itself to a more personal discussion when speaking of sensitive information. Some people are put off by a computer being in the room as it can represent easy access by many people or that “big brother” (government) can potentially access the information now or in the future. For someone that may have a criminal record, a serious mental health condition, or substance abuse history, that idea can impede sharing accurate information. Paper can feel more personal.

The downside to first collecting information on paper is that there is an added step (and staff time) for entering the data the computer. Errors can also be introduced in the process of transcribing the data, and this factor can be increased if intake workers have poor handwriting. On the other hand, the extra step does afford a chance to check information and enter at a slower pace when the consumer is not with you.

Tip: If entering data directly on the computer, consider allowing clients to see the screen as you type or view a report of their information at the end of the interview. This builds trust and enhances accuracy.

Data entry will be much easier if the paper form looks very similar to the computer screen. However, if the computer intake process is not straightforward, it may not make sense to replicate that on paper.

The advantage to entering data directly into the computer is that data entry is done immediately. However, trying to maintain a flow in conversation, while typing, and switching screens leaves room for data entry error and can set an impersonal feel to the interview. The physical presence of a computer placed between the intake worker and staff can also negatively impact rapport. Consider two things if circumstances permit: (1) allow the consumer to see the screen with you as you enter and (2) go back after the consumer has left (immediately if possible) to review that the data entered were accurate. Seeing the screen together shows the consumer you are entering what s/he says and allows him or her to catch a mistake. Alternatively, if the software allows,

Tip: Paper forms should closely resemble the layout of computer screen. Questions should appear in the same order. The paper form should provide checklists of response options wherever possible, and options should match the options in the HMIS.

the intake worker can print a report of the client’s information and present it to the client for review at the end of the interview.

Ultimately the choice to enter directly into the HMIS will depend on whether (1) the HMIS is easy enough and fast enough to be used in real time, (2) the front-line staff is comfortable enough with the system so that it is not a distraction, (3) most of the clients served will

information such as age, race, family size, income sources, disability, veteran status, etc. and then screen the information to determine if the client is eligible for state and/or federal benefit programs. Some software will then generate the application form. Benefits screening requires accurate information otherwise the exercise is moot, so front-line staff and consumers are more likely to collect and immediately enter complete and accurate client data. Other consumer benefits associated with immediate entry of accurate information include: getting accepted into a service program, qualifying for special support within the agency, or not having to complete assessment surveys more than once within the continuum if the data is shared with appropriate agencies.

Data Entry Staff

If data is collected on paper, it must be subsequently entered into the computer. What follows are some key considerations in this process to further ensure data quality.

Data Entry Accuracy

Data entry staff is responsible for entering accurate data. There are a number of unintentional errors that can occur during data entry,

The classic data entry errors are typographical. Such errors can be based on missed keys or transposition of characters. This problem is reduced to the extent that drop down boxes, check boxes, auto-fill, and other tools are used in place of free form text. However, errors are also possible with these fields. One common error is accidentally selecting the wrong response from a drop-down list.

Another type of error is swapped fields, such as entering the last name in the first or middle name field, or intake date in exit date field. The data entry staff person should be cognizant of the layout of the screen and make mental note of any irregularities, such as a form where the last name appears before the first name.

Tip: Data entry staff can catch many errors by proofreading a hard copy report of the data they entered. Different staff members can check each other's work.

Misspelling is another type of error. While the proper spelling should have been recorded by the intake worker, the entry worker should make sure to read the intake form carefully. If the data entry staff is doubtful about the spelling, they should make a note of it and check with the person who wrote it originally. The same is true for questions regarding illegible writing or ambiguous shorthand.

Proofreading

The main way to mitigate the risk of data entry errors is to proofread the data against the original form. It is best to proofread a hard copy. Instead of printing the actual screen and proofreading one client at a time, data entry staff can print a report of all the data on all the clients they entered that day and proofread the report. They can then go back and fix errors after all of the proofreading is complete. If there are multiple data entry people on staff, different staff members should check each other's work.

Professional proofreaders often proofread backwards checking one letter at a time against the original document. This technique can be helpful in checking the free text fields. Reading backwards prevents the mind from seeing what it expects instead of what is actually typed. Reading out loud is another tip. It allows multiple senses to be engaged in the work. Sometimes the ears can catch what the eyes miss.

Another technique is to proof for different types of errors separately. For example, given the types of errors listed in the previous section, it makes sense to first look for misspellings or typographical errors, then for incorrect drop down answers, then for swapped fields. Keeping a running list of the types of errors found can help provide ideas of mistakes to look for in the future.¹³

Training

Standardized training (throughout a CoC) is vital to quality data entry. Software training should be done using a standardized curriculum, presented consistently in computer labs. If a continuum cannot access these resources, it is important for all trainers to use the same training approach and materials, and communicate with each other to assure consistent training.

User training should also cover how to collect data, how to pass data from front-line staff to data entry staff; how to log questions about the data and how to resolve those questions; how to give feedback; and expectations for participating in user meetings. Some of these issues may be program-specific, so they may need to be addressed by internal training rather than as part of the system-wide software training. However, the system-wide trainer might build in "break out" time for individuals to learn their own program's processes.

Who Should Do Data Entry?

Ideally, the same person who collects HMIS data should enter those data into the HMIS. This assures consistent interpretation of the questions, the answers, and handwriting. At many service agencies having one person do

Tip: Intake and data entry staff should meet regularly to resolve any confusion over notes on the intake form, agree on shorthand, and clarify confusing questions. A data quality log can track open questions.

¹³ Adapted from L.R. Communications Systems, Inc., "Proofreading and Editing Tips" http://www.lrcm.com/tips/proofreading_editing.htm (Last visited 3/10/05).

both is not possible; e.g., day shifts might collect the data, night staff might enter it when things are less hectic. Also, the same people who are good at interviewing clients may not be good at entering data, or vice versa.

When it is not possible to have the same person collect and enter data, a clear process and communication between data intake and entry staff is essential. This will minimize any misinterpretations. Staff members doing these two tasks should meet before they begin and consistently check-in to resolve any confusion over notes on the intake form, agree on shorthand usage, clarify confusing questions, and discuss anything else that comes up. Supervisors should ensure that this communication happens regularly at each agency.

Feedback Loop between Data Entry and Intake

Finding out three months down the road that data entry staff were skipping some fields, or interpreting a question incorrectly, can render a period of data useless. Worse would be to never find out about incorrect data entry and use these invalid data in aggregate reporting. This can be prevented fairly easily with a regular feedback loop. A feedback loop simply means building in a regular meeting time to review and answer questions that data entry staff may have for the front-line staff (people filling in the paper forms), and correcting any mistakes and/or misunderstandings before they are repeated multiple times.

At an initial meeting, include the HMIS lead for your agency, all data entry staff and volunteers, all front-line staff, and (ideally) the agency director. Lay out the need for the data, the importance of each role, the meeting schedule, the Data Quality log, the process to resolve any questions about processing the data, and the feedback loop. A **Data Quality Log** tracks information about unresolved data entry issues, such as the date of issue, nature of issue/specific reference, date of resolution and nature of resolution. Data Quality Logs should be part of regular meetings with data entry and front-line staff. (See the sample Data Quality Log in Appendix 4 and the Feedback Loop Flow Chart in Appendix 5.)

Tip: Data handling processes should include procedures for entering new clients, updating existing client information, handling exit data, and re-enrolling returning clients.

Volunteer Issues Regarding Data Entry

Many homeless service agencies don't have the resources to cover all of their HMIS needs and often rely on volunteers for data entry. Volunteers can pose challenges given the fact that turnover is high, and there is little binding them to the position other than their personal commitment and dedication to helping the agency. Volunteers should receive the same training as regular staff and have the opportunity of regular check-in with data collectors (just as regular staff that enter data do). They should be encouraged to log all questions in a data entry issues log and be encouraged to list anything that is unclear. It may be overwhelming at first, but will assure that they have a shared understanding of the importance of their job, and assure the data they enter reflects what the front-line staff and consumer intended.

Agency and Program Directors

Agency and program directors set the stage and maintain momentum in maintaining data quality. They may or may not have a hands-on role, but their management of the process and emphasis on quality data is key.

Executive Directors of agencies set the tone for the organization and play direct and indirect roles. *Direct* roles include actually monitoring data quality processes and tools through regular, substantive meetings with program directors and/or other key staff. *Indirect* roles include keeping data quality “on the radar” and establishing a process to advance data quality goals.

Establishing Processes

Program directors, in consultation with intake and entry staff, establish the workflow processes for gathering and entering data. These processes should include procedures not only for entering new clients, but also communicating information about when the client exits and when data need to be updated. This is especially true when data entry is not done by the person who interacts with the client. Passing information between staff about new clients may be as straightforward as placing the client file in the inbox of the data entry staff. There may be a different process to pass information about exiting clients. The data entry staff may just get a list every day of all the clients who exited with their destinations and other exit information. Or, it may be the job of the caseworker to type in exit dates and other exit information, even if someone else is doing the initial data entry. There might be a third process for handling any other updated information about the client. For example, there may be a separate form that is used by staff to record updates, which could be distributed to data entry staff for entry and subsequently routed to the client’s paper file.

Organizational Support

Data entry in the short-term does not save or hurt lives. In a crisis environment it is extremely challenging to convince users to take the time to carefully enter data. Therefore there must be top to bottom organizational support for quality data collection and entry. If issues related to data collection and entry are never discussed at full staff meetings or in written messages from the director, the impression is given that data entry is not valued as much as other agency work. The agency’s culture should reflect the importance of and commitment to quality data. (See an example of a Director’s Memo in the Appendix 6.)

Tip: Data quality procedures and updates should be folded into already scheduled regular staff meetings.

Data Quality Plans and Job Performance

A program director should create a data quality plan for the program. Data quality plans set benchmarks for data quality, establish monitoring procedures, and incentives for compliance. If there is already a system-wide plan in place for data quality, the director will still need to create internal procedures to meet or exceed the threshold specified by the system plan. In the absence of a community-wide comprehensive plan, program directors should establish their own plan.

Whether the plan is written by the program or the implementation, the director is responsible for ensuring the content of the plan is understood and the benchmarks are achieved by all users. Data quality plans are discussed in more detail in Chapter 2.

If the agency conducts periodic job performance reviews, directors may address data quality as part of that process. For staff directly involved with processing data, the director might link successful completion of tasks (e.g., timely entry, completeness, accuracy) to job performance reviews. This is another concrete way to show that data quality is important to the director and the agency.

Monitoring Data Quality

At most homeless service agencies (especially emergency shelters) finding any extra time to monitor data is nearly impossible. Therefore monitoring data quality should be integrated into the daily flow of running the organization. At regular staff meetings, agency and/or program managers should emphasize the importance of the data, any upcoming needs of the data, and efficient uses of the data within or outside of the agency. For example, a program director can mention that HMIS data for a recent quarter was particularly helpful in completing a grant proposal or that a report of homeless people coming from within the state was used to inform a bill being proposed by state legislators. Seeing the usefulness of the data in real, impact terms is likely to keep staff committed to the process of data quality.

Tip: If a train-the-trainer model is used, the agency's on-site trainer should be allowed time for fulfilling the responsibilities and the role should be built into his or her job description.

If data entry staff keep logs and maintain a feedback loop with front-line staff, the results of open (unresolved) or closed (resolved) issues should be shared semi-regularly as well. Your agency may have other opportunities that better lend themselves to regular check in. What is important is that assuring data quality becomes part of your agency's culture.

Good Training for Staff HMIS Users

Many CoCs use a train-the-trainer approach for HMIS software that delegates training to the agency-level. The benefit to this model is cost efficiency. The downside is the potential that the training won't be as thorough or consistent across agencies or programs. That said, the train-the-trainer model works best with standardized curriculum and materials. Curriculum sounds rather formal, but it is simply a documented approach to what is emphasized in the training and how it is covered. The train-the-trainer approach can be utilized for both software training as well as data quality assurance training. If a train-the-trainer model is used, the agency or program director should ensure that the agency's trainer is supported in this role by having it built into his or her job description and allowed time to fulfill the responsibilities. (Sample user data quality training is included in Appendix 1.)

Mandating Refresher Training for Staff

Refresher training in HMIS software is needed periodically for data entry staff to ensure ongoing data quality. The need can vary depending on the number of changes/upgrades to the software and the overall complexity of the software. It also depends on the skills of the users. Staff that are less comfortable with computers in general should consider refresher trainings to catch mistakes they may be making, and affirm correct usage. All staff can benefit from trainings that go deeper into software. Refresher training in software can vary from every six months to every three years depending on the variables above and the rate and type of data errors.

Use of Data for Program Purposes

HMIS should help staff do its job better, not create new jobs. For example, HMIS can dramatically improve how agency staff assign beds, organize case management, determine appropriate referrals, assess clients' needs, track progress and analyze a program's or an agency's progress in meeting its goals. The more staff and clients benefit from the HMIS, the more data quality will improve. This is not always easy to accomplish, but with that emphasis HMIS can be a support rather than an obstacle, and data quality will benefit.

Tip: Use of client data from HMIS—instead of handwritten notes—in case management discussions ensures that information is in a standard format and requires that data are accurate. Sharing of the files also serves as a data quality check.

Another example is incorporating the use of client data from HMIS in case management or staff discussions. While following your security procedures (e.g., not printing clients' names on meeting documents), sharing these records will ensure that printed client files or reports are in a clear, easy to read, standardized format to facilitate discussion of a client's needs. This feature requires that the data are entered carefully and accurately. If there are data entry errors, the meeting and sharing of the file can serve as a data quality check. Mistakes are more likely to be caught and corrected with more eyes reviewing.

The program director can use the data reporting features to regularly mine the HMIS data for program statistics. These are useful not only for grant writing, funding reports, and advocacy purposes, but also for generally keeping abreast of the number of people an agency is serving at particular times, client characteristics and needs, and what services clients are receiving. If front-line and data entry staff know that directors rely on HMIS data on a regular basis to learn what is happening in the program, data quality is bound to be higher.

Chapter 4: Implementation and Continuum of Care Staff

This chapter describes specific strategies that an HMIS project manager and CoC-level staff can do to foster data quality.

Mechanisms Prior to Entering Data

HMIS project staff can provide all agencies and all data entry staff with good software documentation including a data dictionary and cheat sheets for entering data. It is also important for the CoC to provide consistent and continual training of staff involved in data collection and entry. The CoC should provide extensive training materials if agencies are responsible for staff training and/or consider providing centralized user training. (See training section in previous chapter. See also Appendix 1 - “HMIS Data Quality Training Template” for the backbone of basic data quality user training.)

Data Quality Plans

HMIS project staff members are responsible for developing and enforcing a data quality plan. As previously mentioned, data quality plans set benchmarks for data quality, establish monitoring procedures, and incentives for compliance. Data quality plans are discussed in detail in Chapter 2.

Consistency Among Agencies

The CoC should ensure consistent data collection and quality across all of its programs. This can be achieved through some or all of the following mechanisms:

- **Establishing a continuum-level subcommittee on data quality.** A data quality subcommittee can be charged with making sure data quality remains prominent in Continuum decision-making. Each of the following actions might be implemented and overseen by this subcommittee with frequent reporting to the wider HMIS committee.
- **Convening regular user meetings.** These meetings fulfill many needs. They keep HMIS users/overseers abreast of HMIS efforts across the Continuum. This helps maintain momentum, identify user concerns and software needs, share solutions to common problems and best practices, and provides opportunities to review and refine data quality processes. This is a good use of time by HMIS staff saving them help calls and individual site visits.
- **Conduct routine analyses/comparisons between programs.** Comparisons among CoC programs can serve as a healthy competition to meet the standards the CoC agrees to. It can also serve to identify best practices in data quality and general usage.
- **Defining parameters for data definitions.** The CoC is uniquely positioned to ensure common parameters (or meaning) to questions in the HMIS software. For example, is asthma a physical disability? Is PTSD a mental illness or a separate category? If there is confusion around questions that the HMIS administrator or software documentation cannot

answer easily, the Data Quality subcommittee can discuss and agree upon a convention. This information should be shared throughout the CoC (and also with the software provider).

- **Requiring monthly or quarterly reports generated out of HMIS to verify timely data entry and quality.** For CoCs receiving funding beyond HUD, reporting regularly to these other funders (or just internally) is a way to galvanize agencies and promote a culture where data collection and quality is taken seriously and completed. Going back six months later to catch up on data entry is a recipe for poor data.
- **Programming queries and generating regular data quality reports.** The HMIS project staff can play an important role by providing agencies with standard queries or tools to help them verify their agency's data quality. Similarly, these reports can be run on the overall system data to identify data errors. Sample Data Quality Reports can be found in Appendix 8.
- **Institutionalizing a feedback loop to agencies.** HMIS project staff (or members of the Data Subcommittee) may create a process by which agencies submit data quality updates (examples of data entry issues log, meeting minutes, and reports of data). The subcommittee can use this information to establish a reasonable standard among agencies and help the CoC assess itself on the quality of its data.

Validating and Cleaning Data

Checking data on homeless persons from multiple programs and various ways of entering data is a constant challenge. But, once the data has been collected, there are ways to “clean” the data, that is, fix any errors.

Some CoCs require agencies to print and certify the accuracy of their data (by signing the monthly or quarterly report) as an effort to make them aware and accountable for the information. This can be done through staff discussions about what the aggregate data says. For example, does it ring true that we served 14 people that speak a language other than English? Did we really refer 35% of our clients to intensive case management? Did any of our clients report income over \$100,000 last year? If not, this may be a data entry error.

Tip: A data quality subcommittee can take the lead in convening user meetings, compare quality among programs, and establishing consistent parameters on the meaning of answers.

Agency or program data can be compared with findings from a study by local researchers where there was some overlap in focus. For example, did the local annual census count find 40% families among homeless people in the community; whereas you are finding 20% in your data? What might account for the discrepancy? The census could be wrong, the HMIS data could be wrong, or the parameters could be incorrectly defined. Maybe a large agency serving homeless families has lost their IT staff and data was not entered for the past three months. Validating and cleaning data should also occur at the client level within the database. These can be automatic, or if the software does not check for incorrect data, the HMIS database administrator can do it manually. Some incorrect fields are more obvious than others. Since

validation is best done by the software, a complete discussion of the types of validation can be found in Chapter 5.

At the Continuum-level there are also data validation and cleaning tasks to consider. The Continuum (either through the HMIS system administrator, Project Manager, or CoC Data Committee) needs to establish clear guidelines for agencies across the Continuum. Consider the following:

- **Establishing conventions for dealing with missing data.** For example, this may include generating an automatic program exit date for clients that have not interacted with the program for a certain period of time.
- **Outliers.** Outliers are data that lie outside of an accepted normal range of values. Outliers should be flagged and reviewed by whomever is analyzing the data. Upon review, some outliers may be determined appropriate, some deleted. For example, if monthly income for everyone in the system falls between \$0-\$3,500 except for one “outlier” of \$15,000, you may determine to delete that outlier or double check with the client or case worker. It is an educated judgment call. It may be a yearly income that was incorrectly listed as monthly.
- **Comparing self-reported vs. system generated data.** An example of validating self-reported data against system data is comparing the percentage of people who reported that they stayed in another emergency shelter prior to program entry with the actual percentage of people in the system who were recorded in the HMIS at two or more shelters. If all or most emergency shelters are participating in HMIS, and 60% of clients said they spent the previous night in shelter, but only 10% were recorded in more than one shelter, then it is possible that many clients are not being entered, or something is wrong with either the self reporting process or the data matching process across shelters.

Releasing Data

The Continuum is also responsible for releasing community-wide data as appropriate. A discussion of the release of data can be found in Chapter 7.

Chapter 5: HMIS Software Solutions

The HMIS software is the final link in ensuring data quality. Different software solutions include various features to improve quality. Some of the functionality is customizable at the agency or implementation level. In some cases, the features exist, but may require advanced skills or additional time and effort to use. However, some of the system functionality must be built into the software by the solution provider. Nevertheless, even if particular functionality does not currently exist in a particular system, providers might add features to their products if requested by their clients. This section explores the various ways that HMIS software can be used to improve data quality. In some cases there may be tradeoffs in employing particular features.

Screen Design

Aside from any special functionality, data quality can be greatly enhanced through effective design. Many HMIS solutions include custom screen design options that allow System Administrators to tailor screens and modules for their needs.

It is usually advisable to have all required questions on one or two screens. Software that requires the user to jump between multiple screens in order to capture all required information is likely to lead to many fields that are missed. However, screens that are too cluttered can cause problems with “swapped fields,” in which a user clicks into the wrong field to enter data. Swapping fields is also a problem when the order of questions on the screen is not logical or does not conform to the order in which questions are asked on a paper form or during the interview process. The first name and last name fields are often accidentally swapped for one or more of these reasons, and this problem is very difficult to check for. A related problem is entering the Last Name in the Middle Name field, since Middle Name is often left out.

A good alternative to placing all the fields on one screen that overcomes clutter is a “wizard” intake process that places fields on several different screens, but guides the user through the intake screens with very clear “Back” and “Next” buttons culminating in a completed record.

Tip: Adding an asterisk to the name of a required field is often simple to do using customization features.

Field Labels

Another common screen design technique is to clearly mark required fields. Required fields can be highlighted using boldface, an asterisk, or a different color.

The use of easily understandable field labels is also important for both ensuring that the client provides accurate information and that data has consistent meaning across sites. The *HMIS Final Notice* does not mandate the actual label for the required fields (though the response categories are mandatory) and allows communities to tailor the wording of questions.¹⁴ Thus, screen designers may wish to label the field for “Residence *Prior* to Program Entry” as “Residence *Night Before* Program Entry.” However, labeling the field “Residence Last Night” may lead to poor data when the intake interview is not conducted on the day the client arrives.

Tip: If default responses are customizable by program, consider using them for fields when nearly all clients give the same answer, for example, gender at a women’s shelter, disability status at a mental health facility, or veteran status at a veteran’s shelter.

Some HMIS solutions also provide additional functionality that allows users to access in depth descriptions of the questions and clarifications of the response categories. This additional information can include pop-up text when the user moves the mouse over the word. Or, it can consist of a separate screen that the user clicks to access.

Response Options

For the HUD Universal and Program-Specific fields, the wording and list of response options are mandated.¹⁵ Nevertheless, the ways in which the options are implemented are important. Some situations may warrant using default options for particular question. This is especially true if default options are customizable for particular programs. For example, a men’s shelter could customize the software so that all of the client records default to male. A veteran’s shelter could default to recording that the client is a veteran. If the default option is not used, data entry staff is apt to stop filling in it since they know all of their clients will respond a certain way based on their program eligibility requirements. Default options can help avoid missing fields when nearly 100% of clients give the same answer. Overuse of default options, for example, using a default of “White” when 75% of clients are white is not recommended, since this is very likely to lead to non-white clients being recorded as white.

Another important principle is distinguishing between null and negative responses. Screens that implement “Yes/No” type questions using checkboxes often have this problem. See the example below:

Physical Disability:

This manner of tracking the disability question is not HUD compliant, because it does not specifically state the options “Yes” and “No.” It also leads to poor data quality. If the answer is not checked, it is not possible to determine whether (a) the question was asked and the answer was “No,” (b) the client refused to answer, or (c) the question was never asked at all.

Another issue with response options is the way in which they are presented. In the example below, there is some possibility that users will check “No” when they mean “Yes,” since they

¹⁴ HMIS Final Notice, Section 2 (Introduction).

¹⁵ HMIS Final Notice, Section 2 (Introduction).

need to concentrate and observe that the radio buttons precede the question. The case below is a particularly egregious though not uncommon design, in which the button for “No” is actually closer to the word “Yes”.

Physical Disability: Yes No

A better design is to place the options on different lines:

Physical Disability: Yes
 No

A drop down box can also be used and is preferred when there are numerous response options and only one answer is allowed. Free text boxes should be avoided whenever possible. For example, the “Zip Code of Last Permanent Address” field could use drop down boxes for the common zip codes in the implementing region, and a free form “Other Zip Code” box could be used for infrequent other zip codes. Both fields could populate the same column in the database.

Although communities may have limited control over some of these design issues after software is selected, requests to improve the design for optimizing data quality can be made for software upgrades. In addition, warnings regarding some potential pitfalls inherent in the current version can be included in the user training.

Validation

The feature most closely associated with data quality is validation. The HMIS software can validate that all required fields are entered and that the data meets certain criteria. Depending on the circumstances, the software can either disallow saving the record with missing or possibly invalid data, or it can issue a warning. The stricter approach of preventing the record from being saved when fields are missing raises the response rate but could lead to staff entering false information in order to get past the screen. Alternatively staff may not enter the known information in a timely manner since they would need to wait until a subsequent interview to complete the entire screen.

Types of Validation

- **Field is Null.** Required data is missing.
- **Data type.** E.g., date, number, dollar, or character values.
- **Length.** E.g., complete social security numbers should always be 9 digits long.
- **Logical content.** E.g., date of birth too early or too late.
- **Bogus values.** E.g., SSN equal to “999-99-9999”
- **Conflicting values.** E.g., veterans who are minors.

It is not advisable to equate all the universal fields required by the *HMIS Final Notice* with “required fields” on an intake screen. HUD’s required fields refer to the fields that must be collected by HMIS implementers. However, if a particular client is missing a particular field, such as race, it is better to record the data available than nothing at all. This is particularly true with the “Disabling Condition” field, since disabling condition must be collected separately from

the intake process in order to comply with Fair Housing laws.¹⁶ Use of the “Don’t Know” response is an alternative, but that is supposed to be used when the client does not know the answer, not when the question has not been asked.

In addition to checking for missing information, software can also validate the values in the field. Validation can be based on data type, field length, and on content. Validation based on data type simply checks that a field expecting a numeric value has only numbers and a date value has only actual dates. Names should not include numbers. This could go a step further and validate, for example, that fields based on currency have only two decimal places. Possible validations based on length include Social Security Number as nine digits (assuming the quality code indicates “complete social security number”) and zip codes should have five or nine digits.

Numerous validations can be done based on content. On the simplest level, it is possible to validate whether the birth date is earlier than the present date and later than, say, 100 years ago. Similarly, Program Entry Date should not be in the future. Income might be checked that it does not exceed a certain amount. Another possible, though rarely implemented, validation can specifically prevent the use of commonly used types of false data entry, for example, the software could prohibit using “123-45-6789” or “999-99-9999” for Social Security Number, or “John Doe” as a first and last name, or “Baby Boy” as a first name.

Many validation possibilities emerge when data are compared. Some possible validations include the following:

- Program Entry Date cannot be later than Birth Date
- Program Entry Date cannot be prior to Exit Date
- Children cannot be veterans
- Men cannot be pregnant
- Those receiving SSDI for themselves should be marked as having a disability

For a more comprehensive list of possible areas of validation, see Appendix 7.

Flags for Missing or Stale Data

Another useful tool is to flag records or fields that require updating or completing. This feature supports the completeness and timeliness standard. It may be implemented in multiple ways. One possibility is to indicate with a particular color or an icon any data that is subject to change, but has not been updated for a given length of time (e.g., one or two months). Fields can be marked with yellow or red to indicate “somewhat old” and “very old” data. A shortcoming with this approach is that the user must actually have the record open to see that the data requires updating.

Another possible approach is a module that allows the user to view all records that require attention. Thus, if a user is halfway done working on a record and needs to logout, they can easily return to an unfinished records module to retrieve that record. Similarly, such a module

¹⁶ *HMIS Final Notice*, Section 2.7

might list all active records that have not been updated for thirty days. If the HMIS associates clients with particular users, this module can be filtered only for clients particular to the staff person. Some tools use similar functionality to inform caseworkers of clients who are due for follow-up interviews.

Non-Typing Technologies

Many data entry errors involve typing. Thus, any automation that can substitute for the keyboard could help improve data quality. These tools are also faster than typing. Swipe card technology and biometrics do not alleviate the need for the initial data entry, but they can help with subsequent entry. When typing is used to search for clients, errors in the search field may lead to the system not finding the client, and the client inadvertently being added as new. Alternatively, even if the proper search criteria are entered, multiple clients might be returned. For example, if there are two clients with the same name, a user might pick the wrong person. Swipe cards, however, are usually keyed directly to the client's identification number, eliminating the possibility of improperly matching the client.

Scanners and voice recognition technology could be used to reduce the initial data entry tasks. However, the reliability of these tools is often as poor as direct data entry, and may be difficult to implement.

Additional features that can reduce typing and therefore improve quality are automatic population of certain fields. This is more advanced than the default fields discussed earlier. For example, entering the zip code can cause the automatic population of city name. This eliminates the problem of using multiple names for the same city. Calculations such as adding sources of income to achieve total income are also good candidates for automation. When entering families it is also possible to pre-populate other family members with data from the first family member entered. Last name, prior residence, zip code of last permanent address, perhaps even race are possible candidates for this automation. However, like the use of default fields, it may also lead to problems, as family members do not always share these elements in common. Some of the validation types mentioned in the previous section can also be automatically entered. For example, the veteran status can be automatically populated as "No" for clients who are minors.

Tip: In the absence of built-in validation and quality reports, simple custom reports can be used to identify clients for whom required fields are null, data are illogical, values conflict, or data needs updating.

Reports and Logs

Reporting can be used as a tool after initial data entry to improve quality.

Single client reports can be printed and used to confirm intake information on the spot with clients if data are being entered in real time. This will also serve the purpose of reassuring clients that only the answers they give are being recorded (assuming this is the case). Note that

proper disposal of these reports are necessary to maintain confidentiality. They should not be left lying around the premises.

There are various types of reports that are designed specifically for data quality.

Client-level data quality can be used to in the absence of functionality to flag missing, invalid and stale data, as discussed in previous sections. For example, canned or custom reports can check for completeness, by listing all clients where the any of the Universal fields are null. A report could also check for conflicting data by listing any client where the exit date is earlier than the entry date. To check for timeliness, a report can list clients where the date of last update is earlier than 30 days ago.

Aggregate data quality reports can provide overall feedback on the quality of the data entered by the CoC, a particular program, or a particular user. These reports can be broken down by field, and a percentage can be generated for the number of records entered in a particular period that completed that field. Thus, for example, a report can inform the continua that a particular program has been entering Social Security Number only 50% of the time. This information can be relayed back to the program as part of the feedback loop. A similar report can be run for each user. These reports might also help to indicate whether the program or user has become lax with entering records in general.

Sample Data Quality Reports are included in Appendix 8.

Standard Aggregate reports are not necessarily intended for data quality; they are the basic output of the HMIS, which might describe the number of clients served, and the frequencies of various characteristics. When program executives view these reports on a regular basis, they can often spot anomalies, such as a decrease in the number of clients entered or a rise in the frequency of a particular characteristic (e.g., the number of clients coming from the same zip code). These anomalies could be investigated, traced back to data quality issues, and corrected.

Such issues are easier to spot with reports on the program level than the CoC level, since the program manager has a good sense of the reality that the report represents. For example, the program manager would know that the shelter has been busy and that a decrease in clients is a data entry problem; whereas the HMIS project manager might assume that a drop from 150 clients to 120 clients is potentially accurate. Also, when added to a large continuum-wide data set issues may escape notice. For example, if all clients in a particular program were erroneously listed as African-American, this might only change the overall percentage of African-Americans in the HMIS by one percent. But the program director would certainly notice if a report falsely listed 100% of clients as African-American.

Some solutions incorporate data quality features into standard reports. For example an HMIS can produce the HUD Annual Progress Report for programs, and as part of the report also indicate whether all the numbers in each section add up to the total number of clients being reported on. If the numbers do not add up, the user can trace the problem back to missing data.

Finally, if data lapses are found, **transaction logs** can identify users who have or have not logged onto the system and who modified particular records. These logs can be used in conjunction

with other reports to help identify staff with persistent problems that could be addressed. Transaction logs can also track when data are entered, which is useful for monitoring timeliness. Reports can be generated comparing the actual date of data entry to the backdated intake date. This can be an effective tool for monitoring timely data entry.

As an alternative for enforcing timeliness, some HMIS software solutions do not allow data to be backdated, thus enforcing data entry at point of service. The program entry date is always recorded as the date the data were entered. Whether this can be considered a data quality feature to enforce timeliness or simply a lack of functionality is debatable. This solution might lead to inaccurate information. This final example illustrates that there are many tradeoffs in implementing data quality features.

Chapter 6: Putting the Pieces Together: A Case Study of Michigan State HMIS

Overview

This chapter looks at a real life example of how the State of Michigan has taken on data quality issues, the processes they have implemented, and the challenges they face. This information is based mostly on a lengthy discussion with the director of the Michigan Statewide Homeless Management Information System (MSHMIS), Barb Ritter, on March 8, 2005, and follow-up correspondence.

In the State of Michigan, the Balance of State Continuum consists of 43 locally defined Continuums. Michigan also has 17 independent Continuums. Fifty-seven of these 60 locally defined continuums are using one HMIS on a single server. The database is owned by the Michigan State Housing Development Authority (MSHDA), which subcontracts with the Michigan Coalition Against Homelessness for database management. MSHMIS is a highly organized network that relies on routine communication to ensure successful privacy protection, implementation, reporting, and data quality. Each of the 57 locally defined Continuums is required to participate in the system (the other three CoCs upload their data regularly for statewide reporting) and each must have its own system administrator. An agency administrator is also appointed from each agency. These roles are essential in maintaining the communication of best practices, requirements, data reports, and problems with the system—all essential to maintaining data quality. MSHMIS is noteworthy in its processes and attention paid to data quality. It also provides a good look at data quality issues at various levels of organization: front-line and data entry users, agencies, Continuums, and statewide. Table 6.1 shows the various levels of MSHMIS organization, including mandatory meeting schedules and quality reporting done at each level.

Table 6.1: MSHMIS Organization

Level	HMIS Leadership	Meetings	Quality Reports
Statewide	MSHMIS staff	Quarterly , mandatory for all System Admins. Multiple meetings held each quarter for CoCs in different implementation stages.	MSHMIS conducts regular system-wide data quality queries and supplies filters for local SAs to use to check agency level data quality.
Continuum	System Administrator	Monthly , mandatory for all Agency Administrators within CoC. MSHMIS prescribes minimum agenda topics. Minutes forwarded to MSHMIS Staff.	Data reports committee in each CoC looks at data quality regularly.
User and Agency	Agency Administrator	Quarterly , mandatory for all users in agency. MSHMIS prescribes minimum agenda topics. Minutes forwarded to System Administrator.	Monthly report required. Expected to run queries validating agency data.

Many of the documents and procedures that Michigan uses can be found online at their website www.mihomeless.org. (An excerpt of Michigan's Policy and Procedure document on Quality Assurance and Data Quality--hereafter "MSHIMS Policy"--can be found in Appendix 9 of this document.)

Philosophy

Ms. Ritter is the MSHMIS Project Director. She brings a zealous approach to ensuring data quality because it ensures people are counted. People that are not counted, she says, "don't exist" in people's minds: "If we didn't count the mentally ill among our homeless there were no mentally ill among our homeless." She says that to sustain data quality in a system, you must do three things:

- (1) Create a dialogue about data that is systematic and routine,
- (2) Monitor the data aggressively and often, and
- (3) Use the data.

Briefly, here is what each involves.

Create a dialogue about data that is systematic and routine

Regular meetings at every level provide a forum to share best practices in entering and monitoring data, and they keep pressure on sites to maintain quality control practices. They also provide support to programs by letting them know they are not alone in their challenges and generally help to maintain momentum. On the technical support side, regular meetings lower the need for calls and on-site visits from program staff.

Monitor the data aggressively and often

The MSHMIS staff regularly monitors the data at the aggregate level and works with program administrators to do the same for their agencies. They look for data outliers and general errors and review findings at monthly meetings of all agency administrators.

Use the data

When Ms. Ritter was doing HMIS implementation in Spokane, WA, she published a "Point in Time" bi-monthly report on the city's website, which included unduplicated counts and other data on homeless and at-risk populations. Her team also developed a protocol for rapid response to a wide variety of requests for aggregated information on specific issues, including service providers, press, students, and policy planners. HMIS data were used in newspaper articles several times a month. Data were also used for funding reports and outcome measurement.

She aims to use the same approach in Michigan and says the key is to make the data useful for agency staff in their day-to-day work. "The degree to which data is useful is the degree to which it is used." When it is useful it is entered more accurately, people pay attention to the data and notice errors, and quality improves.

As an example of how using data can help, if one program staff person continually entered the year of birth as “2004”--the year entered instead of the year born--this error would be noticed at the aggregate level when the reports suggested a huge jump in homeless youth being seen. Ms. Ritter would find this because she regularly sorts the data from top to bottom and could notice a major “jump” or discrepancy from one program to another. Ms. Ritter finds this method of sorting columns to search for outliers particularly effective.

Additionally, aggregated numbers will reveal training failures. For example, a new hire in one particular program didn’t understand what a readmission was. So, she entered all the clients as ‘readmitted.’ This error was obvious when the monthly reports were run. Regular releases allow for regular monitoring and maintain ongoing data quality. In general,

Sending reports without detailed review is dangerous ...because many funding sources take performance numbers seriously. Early in my career I had a couple of experiences where programs [were] actually endangered and in one case lost [its] funding” on account of not reviewing data...When you start using and aggregating that’s when you start catching. Before that people only give it a cursory look. But now that it counts, people find mistakes. Outliers begin to really impact conclusions.

To implement these three principles, MSHMIS employs a bottom-up and top-down approach among staff to catch data entry errors. There is also a carrot and stick approach to ensure buy-in among program staff. The carrot is making the HMIS absolutely useful to staff, directors, and continuum planners through the reports it generates. The stick is that the State of Michigan requires agencies to use the system in order to receive funding. In order to participate, agencies must identify an agency administrator that attends monthly CoC meetings and ensure that users attend user training. A local “Reports/Data Committee” in each of the CoCs reviews the quality of the aggregated data and the conclusions drawn. These requirements are non-negotiable.

Data Quality Standards

The HMIS implementation in Michigan addresses each of the data quality standards (discussed in Chapter 2 of this document). Regarding **timeliness** of data, each agency is required to enter data and is expected to run a *monthly* report that is shared with MSHMIS staff. Data **completeness** is monitored regularly through queries created at the system level and run at the agency level, with constant communication among peers and MSHMIS staff. **Accuracy** and **consistency** are similarly explored through regular meetings and regular running of queries. Inaccuracies are sent back to agencies to fix. Accuracy is greatly enhanced through regular monthly reports, which forces all stakeholders to examine the data closely and catch poor data. All of this is coordinated through data quality plans that are enacted from both the top-down and the bottom-up among an organization’s staff.

User and Agency Level

Agency-level user meetings are required quarterly. For intake and data entry staff, these meetings are the core process by which staff communicates about data quality issues. They are based on the belief that there will be no data quality if people do not talk regularly.

As described in the excerpt from the MSHMIS Policy, the purposes of the user meetings are:

- A. Opportunity for benchmarking between participants:
- B. Review core processes and related measures.
 - (1) Identify issues and share solutions.
 - (2) Identify those issues where additional help is needed.
 - (3) Incorporate process and outcome measures. (For example MSHMIS requires coverage rates types of reports generated at the agency level)
- C. Support transparency.
- D. Share successes.
- E. Review aggregated data.
- F. Formalize communication to database administrators.
- G. Provide routine consumer satisfaction input.

MSHMIS prescribes the agenda topics that must be discussed at each agency-level user meeting. Minutes from those meetings are sent to MSHMIS staff to ensure that a formal dialogue process is occurring with some consistency across the statewide system.

One key to making user meetings work in Ms. Ritter's mind is making the user meetings fun and engaging; for example, she suggests that agencies always provide food. If people dread these meetings they become disengaged, or they don't come. The staff also makes sure that everyone talks at the meetings at some point.

Ms. Ritter usefully distinguishes between systematic and random errors. Systematic errors are concentrated on particular fields within a particular agency and often reflect training issues. For example, users may not be searching properly, or they may misunderstand the meaning of a question. For example, one agency during a particular month left the Date of Birth field blank on five or six records; in another program duplicate records spiked in a particular month. To address these issues MSHMIS staff presents reports to users and inquire about spikes. It is here that the sources of the errors are often discovered through discussion. Random errors, by contrast, are not concentrated on one particular field. However, when many random errors appear in a single agency, this could indicate that staff is not given enough time or quiet space for data entry. Through regular user meetings these problems are more easily identified and remedied. The agency user group is expected to look at their data as a group and make recommendations around specific agency data. MSHMIS cannot publish any data without the agency's permission.

In addition to overseeing the HMIS within the agency, the agency administrator is responsible for sharing information on HMIS and quality assurance to all agency users and is the contact for decisions at the state level, through their CoC-level System Administrator. They also convey information about the status of HMIS at their agency to the System Administrator. One way this is accomplished is by sending the System Administrator structured minutes from the Agency Users Meetings.

Beyond the meetings, agency administrators are also expected to run queries at the agency level to validate their data. They look for errors, such as random missing data for minimally required fields. Through this hands-on familiarity with the data, they can identify any workflow issues that need to be addressed.

Continuum of Care level

The CoC-level system administrators receive and review reports from the agency administrators. They also convene monthly Continuum of Care meetings, which agency administrators are required to attend. These meetings allow an opportunity to create benchmarks, solve issues, and report on monthly data challenges. This creates pressure on each agency to keep their data current, to monitor it, and to “clean” it in time for the monthly reporting “From a data quality perspective, having local eyes engaged in the process makes all the difference in the world,” says Ms. Ritter. As on the agency level, MSHMIS prescribes core agenda topics for each month’s meeting, and the CoC system administrator sends structured monthly reports of these meetings to the MSHMIS staff.

Each CoC must also assemble a data committee from among their Continuum of Care representatives. The CoCs can decide how often this committee meets and how it operates, as long as they look at the data semi-regularly. There is great autonomy in how these committees operate, but from MSHMIS staff emphasizes full integration of the data committee.

Statewide Implementation Level

In addition to receiving monthly reports from the CoCs, MSHMIS staff convenes quarterly cross-CoC system administrator meetings. Not all CoC representatives across the state attend the same meetings. Instead, multiple meetings are held for those in similar stages of implementation. All of the leadership within the HMIS network in Michigan (agency administrators, CoC representatives and data committee members) is represented and the frequent meetings and sharing of information ensures that everyone in the system benefits from all lessons learned.

MSHMIS staff also runs data quality queries developed within the HMIS software on aggregate HMIS data. Any data quality queries are also filtered down for agency administrators to run on their data. Ms. Ritter: “I teach that these databases are good capturers but not good analyzers. People have to use their data. If it’s always done at the top, no one at lower level learns or can run data.” Queries that MSHMIS staff have developed to check agency-level data quality look for the following:

- A. Null DOB and gender fields
- B. Rate of infants under the age of 1
- C. Gender by family relationship
- D. Homeless by “extent of homelessness”
- E. HUD Assessment by entries & exits
- F. Age by family relationship
- G. Number of users, CoCs, and records on the live site

- H. Null exit dates related to short term services
- I. Ambiguous data in reports.¹⁷

Ideally, the Michigan model works from the ground up with MSHMIS staff serving as consultants and facilitators to the CoC System Administrators and through routine training of program staff. In areas with fewer HMIS resources, MSHMIS staff also does more direct work in the community.

Most of what we do is people time and skill sets. How do you ever get to data without a process around it? We don't hire statisticians, asking them to look at and aggregate data. It's the same process for an agency, for CoC level, and for the state level. I set up trainings for the database and serve as a consultant, we don't take over decision making, so the implementation looks a little different in each CoC.

Having cash incentives helps. For CoCs without dedicated HMIS projects, MSHMIS provides a stipend (between \$2,500 and \$4,000) to update computers, purchase high-speed access, or fund IT support. Each CoC uses their grant to address whatever local barriers they identify.

“The rest,” says Ms. Ritter, “is social will.”

People value quality data because it allows them to respond to their program's need for information, to make a local case for those who are poor, to improve the quality and the ease of their measurement, and get unduplicated counts. Our primary objective is to help them use their data to address first their program needs and then their community needs.

Challenges

Like all implementations, Michigan has challenges when it comes to data quality. These challenges include:

- Discharge issues,
- Unduplication, and
- Data or definition drift

Discharge Issues

Discharging or exiting a client from an HMIS system is difficult, especially for homeless people that move in and out of the service system. In many ways, the current framework for thinking of clients entering and exiting does not work, particularly in emergency shelters. Many clients come and go as the need for shelter arises. If someone has a pattern of staying in the shelters for a couple of weeks then leaving the area for a month, then returning for a day, then gone again indefinitely, at what point do you consider them discharged? Some programs around the country are defining a client as discharged/exited after a period of 30 days have passed. When he or returns the return is counted as another episode.

¹⁷ MSHMIS Policy

Determining how the community defines and documents a client's discharge or exit from the system has implications for overall numbers. If a funder is going to penalize a program for accurately calculating when a client has left the program there is no incentive for programs to change how they record discharge. MSHMIS advocates that the community should come together to reach a shared understanding. Funders need to understand that if service numbers go down due to a change in definition that does not necessarily mean the program is underachieving.

Unduplication Algorithm

An unduplication algorithm is a mathematical formula the HMIS software uses to determine a unique client I.D., search for a duplicate of that client I.D., and count one instance of each unique I.D. This algorithm depends upon correct name spelling, DOB, and gender. If any of those fields are entered incorrectly the unduplicated count is wrong.

Beyond the agency level where unduplication is less of an issue, analysis requires the use of secondary databases that are created for generating statistics. The first step in that process is to remove the obvious duplicate clients. Another challenge is to maintain a process for going back and correcting the master database.

Data or Definition Drift

Data drift describes when an agency's data slowly starts deteriorating and showing missing data or incorrect data. It is usually at one agency, not system-wide, and can usually be recognized by running data totals.

Missing data is easy to spot, but how do you determine when data are not good? Data can be entered incorrectly, or questions can be asked incorrectly, or misinterpreted. For example, take the questions, "How many times have you been homeless?" "When was the last time you were housed?" In those cases where different meanings are assigned to the question, MSHMIS staff can determine from user meetings where people may have generally been coming from and what answers are outliers. Questions like this are difficult, if not impossible, to answer in a bullet or single drop down choice. Staff also finds that people often answer different ways at different points in time. Absent a controlled quality interview process, housing history is especially vulnerable to problems with stable definitions. MSHMIS staff struggle over what to do with such data.

The procedures associated with data collection also vary widely from a detailed interview by highly qualified staff to handing the client a clipboard and asking them to complete a few questions. The selection of data to include in analysis must be informed by an understanding of the local collection process and the types of information being aggregated. Simple prevalence counts can include everyone, but data on special needs should only be gathered from those programs that involve relationship building with the client.

To meet these challenges, MSHMIS takes issues and ideas back to the user meetings. Keeping up the education of users is essential, talking about these issues and sharing what others are finding tends to lead to a solution.

Conclusion

For the Michigan State HMIS project, everything rests on data quality. They have one of the most ambitious projects in the country.. Their adherence to regular communication, monitoring the data, and ultimately using their data regularly is what they feel will ultimately make their system succeed. The MSHMIS philosophy is that all efforts at maintaining data quality must come within the flow of the workday (not added on as separate task for its own sake) and ultimately the data generated must help all staff in the process.

“I am not a researcher; I am a performance enhancer,” says Ms. Ritter. As a performance enhancer, she is banking on the fact that performance of the system is equal to the quality of data. As one goes, so goes the other.

Chapter 7: Releasing Data

The goal of HMIS data is to represent the individuals and families who access homeless services at the program level, CoC level, statewide and nationally. The most important product of all the efforts aimed at improving data quality is to release information that accurately represents the homeless population. Even where data quality falls short, actively monitoring the quality can help to understand and describe the scope and limitations of any data released. If a report does not accurately describe the whole population, it should at least be able to accurately represent its own limitations. Whether or not you have adequate coverage, any data reports should be accompanied by very clear parameters of how the data was collected and what the data can and cannot claim. (See Coverage Rate and Parameters Detail in Appendix 10.)

To achieve the goal of accurate representation, close attention to which programs have contributed data to the HMIS and the quality of HMIS data is critical before any data release. Potential pitfalls of releasing low-quality data include producing invalid information leading to program and policy redesign that does not meet the needs of homeless people. It may also lead to an unduplicated count that is either too high or too low, which sets a false benchmark for analyzing future trends.

Overall Coverage

Adequate client coverage is needed to claim that data fairly represents of your program, community, or implementation. Determining an adequate level coverage is not straightforward, and various models exist. For the first Annual Homeless Assessment Report (AHAR), bed coverage is calculated by dividing the total number of beds offered by providers participating in HMIS by the total number of beds in the community. The threshold for minimal bed coverage using this formula was determined to be 50% for AHAR site and 75% for other contributing sites, although a 50% coverage level is acknowledged to be less accurate than higher coverage. This calculation is done separately for individual emergency beds, individual transitional beds, family emergency beds, and family transitional beds.¹⁸

The initial AHAR coverage guidance is a bare minimum threshold and communities may determine that caution requires a stricter standard, before reporting on the data. The stricter standard may involve a higher percentage (e.g., 60% or 75%), but it can also involve a more precise way of calculating coverage. For example, rather than simply calculating the beds in participating programs versus total beds, it is possible to incorporate into the calculation the extent to which full capacity is being reached in each of the various shelters, whether turnover rates vary between shelters, and whether all the clients entering the shelter are being recorded in the HMIS.

Any statement regarding the representativeness of the report should clearly state the scope of the report in terms of the types of programs analyzed. For example, if the data only covers unaccompanied individuals in emergency shelters, it should be very clear that families, those using transitional or specialized shelters and those who stayed on the streets are not included.

¹⁸ “The Annual Homeless Assessment Report: Frequently Asked Questions” (2005).
<http://www.hud.gov/offices/cpd/homeless/hmis/standards/hmisfaq.pdf>

Types of Program

Characteristics of homeless service users can only be generalized to users of the programs contributing data to the HMIS. For example, currently many programs serving domestic violence victims are not contributing data to HMIS due to concerns about protecting their clients' privacy. As such, data released on homeless shelter clients may report fewer homeless women or families than the number actually served and will under report rates of domestic violence.

Similarly, if the veterans shelter is not participating in HMIS, this missing information might greatly affect not only the number of veterans, but also the gender ratio. Skews such as these should be clearly indicated and highlighted in any data findings so the data is not misrepresented or misused as representative of your entire Continuum.

Data Fields and Sub-Groups

The report should state limitations caused by incomplete data records. For instance, a report cannot accurately state the frequency of client characteristics, such as the percentage of people who are Hispanic/Latino, if the primary data field is null (empty) for a significant portion of the client records. Similarly, it is not necessarily legitimate to calculate the average age if only a small fraction of those clients supplied their date of birth. Since there will inevitably be some level of incompleteness, it is important when reporting on these frequencies to indicate the universe of records that actually answered the question being reported on. This will enable readers to better understand the level of validity of a finding. If the universe of respondents for a particular field drops too far below the total universe of records, that data would not be valid. This is particularly true if the overall program coverage rate is also low.

Furthermore, even if the coverage threshold is reached, that does not necessarily mean that one can validly analyze a sub-group or slice of that population. Thus, while it may be valid to report a racial breakdown of the entire population, it may not be valid to report a racial breakdown of elderly women within that same population nor would it necessarily be legitimate to run reports on particular geographic regions, within the overall reporting area.

The CoC should discuss and establish valid policies on levels of field coverage required for data release. The community may also consider establishing a procedure to vet data prior to release.

Tip: The types of programs, in addition to the percentage of programs participating, can limit the usefulness of data. A large veteran's shelter will have a disproportionate number of men, as well as veterans than other shelters.

Validating Data

A number of procedures to check HMIS data are needed to ensure that these data reflect reality as closely as possible. Analyzing data reveals the extent of data entry or data coverage by detecting missing data and incomplete information. Analyzing data also reveals data entry errors. For example, when running frequency distributions, invalid entries or outliers (data that

have extreme values, such as a monthly income of \$3,000) can be detected. (Issues regarding validating and cleaning data were discussed in Chapter 4.)

If available, HMIS data could also be compared to other similar data on the same population for data validity. In sum, data quality is expected to improve with the use of HMIS data. To assist with data quality checks, the optimum approach is to hire an experienced researcher who is trained in social science research and knowledgeable in the field of homelessness and homeless service provision. A researcher can provide support in data analyses and assessment of the data quality before release.

Soliciting Feedback

Before releasing data it serves all stakeholders to release draft findings to the HMIS Governing Committee, service staff, and other stakeholders for review and feedback. Insight provided on a draft report can still be considered as part of final revisions. Diverse perspectives and eyes can catch errors that might be missed otherwise. If the CoC is not diverse enough, you may want to release a draft to other members within the community. Ideally most or all of the following perspectives should have a chance to weigh in:

- Researchers
- Consumer representatives
- Homeless Advocates
- Local Government
- Those serving Special Populations
 - Chronically homeless individuals
 - Homeless families
 - Homeless youth
 - Elderly
 - Domestic violence

Consumers in particular can offer insight into the meaning of certain data that people without experience in the service system, or of being homeless, cannot provide. For example, when the Consumer Advisory Committee reviewed HMIS data on marital status in Massachusetts' individual emergency shelters, one individual noted:

Very few people in shelter who are married will tell you that they are married. They don't want to say, 'I am married with kids and can't take care of my family.'¹⁹

The consumer's comment was released along with the data.

Review prior to release assures that the release is clear in what can/cannot claim, catches any errors, and prepares the Continuum for any ramifications of the findings.

¹⁹ Meschede, T; Sokol, B. ,Raymond, J. et al. (2004) Hard Numbers, Hard Times: Homeless Individuals in Massachusetts Emergency Shelters 1999-2003. University of Massachusetts, Center for Social Policy, 18.

Conclusion

This chapter has focused briefly on the large issues involved in releasing data. It serves mostly as a word of caution regarding the issues that need to be considered when data quality is less than optimal. The problems of releasing data are minimal when data quality is high, but are potentially insurmountable when not enough attention has been paid to data quality throughout the collection process. While cleaning data and accurately representing limitations are valuable, engaging all parties—front-line staff, data entry staff, program directors, implementation-level staff, and the software itself—in a concerted effort to collect high quality data is much more important and ultimately the only way to truly achieve the long-term goals of HMIS.

Appendix 1: Training Template

A PowerPoint “HMIS Data Quality Training Template” was developed in conjunction with this document. It can be downloaded from the HUD HMIS website.

<http://www.hud.gov/offices/cpd/homeless/hmis/index.cfm>.

Appendix 2: Universal Data Elements Data Entry Form

Please note that this data collection sheet should resemble the appearance of the data entry software as closely as possible to reduce the rate of data entry error when transferring information from paper to the HMIS.

FOR FREE TEXT FIELDS, USE BLOCK LETTERS. OTHERWISE, MARK APPROPRIATE BOXES WITH AN “X”

Key: **Q:** Question as suggested by Data Standards,
S: Suggestions on how to probe when clients can't don't know how to respond to the standard question.

Program Entry Date

		/			/				
month	day		yea						

Program Exit Date

		/			/				
month	day		yea						

Current Name (first, middle, last name, suffix)

Q: What is your first, middle, and last name, and suffix (legal names only, avoid aliases or nicknames)

	Don't Know	N/A	Refused
First name	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Middle name	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Last name	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Suffix	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Q: Have you ever received services using any other name?

	Don't Know	N/A	Refused
First name	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Middle name	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Last name	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Suffix	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Social Security Number

Q: What is your Social Security Number?

			-		-				
--	--	--	---	--	---	--	--	--	--

Date of Birth

Q: What is your birth date?

		/			/					
--	--	---	--	--	---	--	--	--	--	--

month day year

(If complete birth date is not know: What is your age?)

--	--

Age

Ethnicity/Race

Q: Are you Hispanic or Latino? (of Cuban, Mexican, Puerto Rica, South or Central American or other Spanish culture of origin)

Non-Hispanic/Latino	<input type="checkbox"/>
Hispanic/Latino	<input type="checkbox"/>

Q: What is your race (you may name **more than one** race)

American Indian or Alaskan Native (origins in any of the original peoples in North, Central, and South America, and who maintains tribal affiliation or community attachment)	<input type="checkbox"/>
Asian (origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent)	<input type="checkbox"/>
Black or African American (origins in any of the black racial groups of Africa)	<input type="checkbox"/>
Native Hawaiian or Other Pacific Islander (origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands)	<input type="checkbox"/>
White (origins in any of the original peoples of Europe, the Middle East, or North Africa)	<input type="checkbox"/>

Gender

Q. Are you male of female?

Male	<input type="checkbox"/>
Female	<input type="checkbox"/>

Veteran Status

Q: Have you ever served on active duty in the Armed Forces of the United States?

No	<input type="checkbox"/>
Yes	<input type="checkbox"/>
Don't know	<input type="checkbox"/>
Refused	<input type="checkbox"/>

Disabling Condition

Q: Do you have a physical, mental, emotional or developmental disability, HIV/AIDS, or a diagnosable substance abuse problem that is expected to be of a long duration and substantially limits your ability to live on your own?

S: If client is not sure, you may want to add: Have you ever been diagnosed with a physical, mental, emotional or developmental disability, HIV/AIDS, or a diagnosable substance abuse problem?

No	<input type="checkbox"/>
Yes	<input type="checkbox"/>
Don't know	<input type="checkbox"/>
Refused	<input type="checkbox"/>

2.8 Residence Prior to Program Entry

Q: Where did you stay last night?

S: If the client stayed in the program last night, ask: Where did you stay on the night before entering the program?

Emergency shelter (including a youth shelter, hotel, motel, campground paid with emergency shelter voucher	<input type="checkbox"/>
Transitional housing for homeless persons (including homeless youth)	<input type="checkbox"/>
Permanent housing for formerly homeless persons (such as SHP, S+C, SRO Mod Rehab)	<input type="checkbox"/>
Psychiatric hospital or other psychiatric facility	<input type="checkbox"/>
Substance abuse treatment facility or detox center	<input type="checkbox"/>
Hospital (non psychiatric)	<input type="checkbox"/>
Jail, prison, juvenile detention facility	<input type="checkbox"/>
Room, apartment, or house that your rent	<input type="checkbox"/>
Apartment or house that you own	<input type="checkbox"/>
Staying or living in a family member's room, apartment, or house	<input type="checkbox"/>
Staying or living in a friend's room, apartment, or house	<input type="checkbox"/>
Hotel/motel paid for without emergency shelter voucher	<input type="checkbox"/>
Foster care home/foster care group home	<input type="checkbox"/>
Places not meant for habitation e.g., (vehicles, abandoned building, bus/train/subway station/airport, or anywhere else outside	<input type="checkbox"/>
Other (Describe)	<input type="checkbox"/>
Don't know	<input type="checkbox"/>
Refused	<input type="checkbox"/>

Q: How long did you stay at that place?

1 week or less	<input type="checkbox"/>
More than 1 week, but less than 1 month	<input type="checkbox"/>
1 to 3 months	<input type="checkbox"/>
More than 3 months but less than 1 year	<input type="checkbox"/>
1 year or longer	<input type="checkbox"/>

2.9 Zip Code of Last Permanent Address

Q: What is the zip code of the apartment, room, or house where you last lived for 90 days or more?

Zip code							
Don't know	<input type="checkbox"/>						
Refused	<input type="checkbox"/>						

S: If zip code unknown, what is the city and state you last lived for 90 days or more?

City:																		
State:																		

Unique Personal Identification Number (PIN)

To facilitate the search for an existing PIN, you may want to ask:

Q: Have you ever been served by this [name or facility or program] before?

No	<input type="checkbox"/>
Yes	<input type="checkbox"/>
Don't know	<input type="checkbox"/>
Refused	<input type="checkbox"/>

Household Identifier Number

If it is not evident that others are applying for or receiving assistance with the client, then you may want to ask:

Q: Is there someone else who is applying for (or receiving) assistance with you? If yes,

Q: What is their first, middle, and last name? (legal names only, avoid aliases or nicknames)

Q: Do you have any children under 18 years of age? If yes,

Q: What is (are) the first middle, and last name(s) of the child(ren) with you?

Please fill out separate form for each family member and clip together.

Appendix 3: Data Quality Plan Work Sheet

This document lists a series of questions to consider answering to ensure a data quality around the four data general quality standards: timeliness, completeness, accuracy, and consistency. It is intended as a guide to creating and implementing benchmarks for your HMIS implementation and/or region in assuring quality data that can help you better understand and address homelessness in your community.

Standard	Benchmark	Questions to Consider	Possible answers**	Decision/Policy
Timeliness	Data are entered soon after collected	How soon should all clients be entered by?	Basic info within 24? Full service, goals, etc within 48 hours?	
	Changing data are kept up to date	How often should active clients be reviewed?	Active clients every 30 days?	
Completeness	All clients entered	What portion of record for all clients?	100% of client identifiers?; All demographic information? Entire record?	
	Complete identifying data entered	What percentage of clients?	95% of clients will have demographic information	
	Complete characteristics fields entered	What percentage of clients?	90% of records will have complete characteristics fields	
	All services entered	What percentage of clients?	85% of clients will have all services entered	
	Complete exit data entered	What percentage of clients?	85% of clients will have exit data entered or an exit date entered for 3 months past last interaction with client	

Standard	Benchmark	Questions to Consider	Possible answers**	Decision/Policy
Accuracy	Truthfulness from clients	How can we assure client's information is truthful?	Tie information given to eligibility benefits; increase trust through data collection process; Show client what is entered; give client a report of record; collect basic info in first 24 hours, and more personal information in 48 hours; explain privacy procedures; offer chance to tell their story not just pieces of data; clarify who uses data	
		How do we measure truthfulness?	Check random sample quarterly (2 per agency) asking client that has established trust to review information originally entered	
	Accurate data entered by staff	How do we assure data is entered correctly?	Use forms that require clear lettering (see example); feedback loop between data entry staff and data collection staff and supervisor; repeat answers and spelling back to client. Proofread.	
		How do we measure accuracy?	Random sample checks of paper vs. computer data every 2 weeks	
		What is accuracy benchmark?	95% accuracy of sample check	
Consistency	Common interpretation of questions and answers	How do we ensure common definitions of fields and how to word questions consistently?	Communication! User group meetings every month; on-line or printed document with detail about minimal elements updated every quarter	

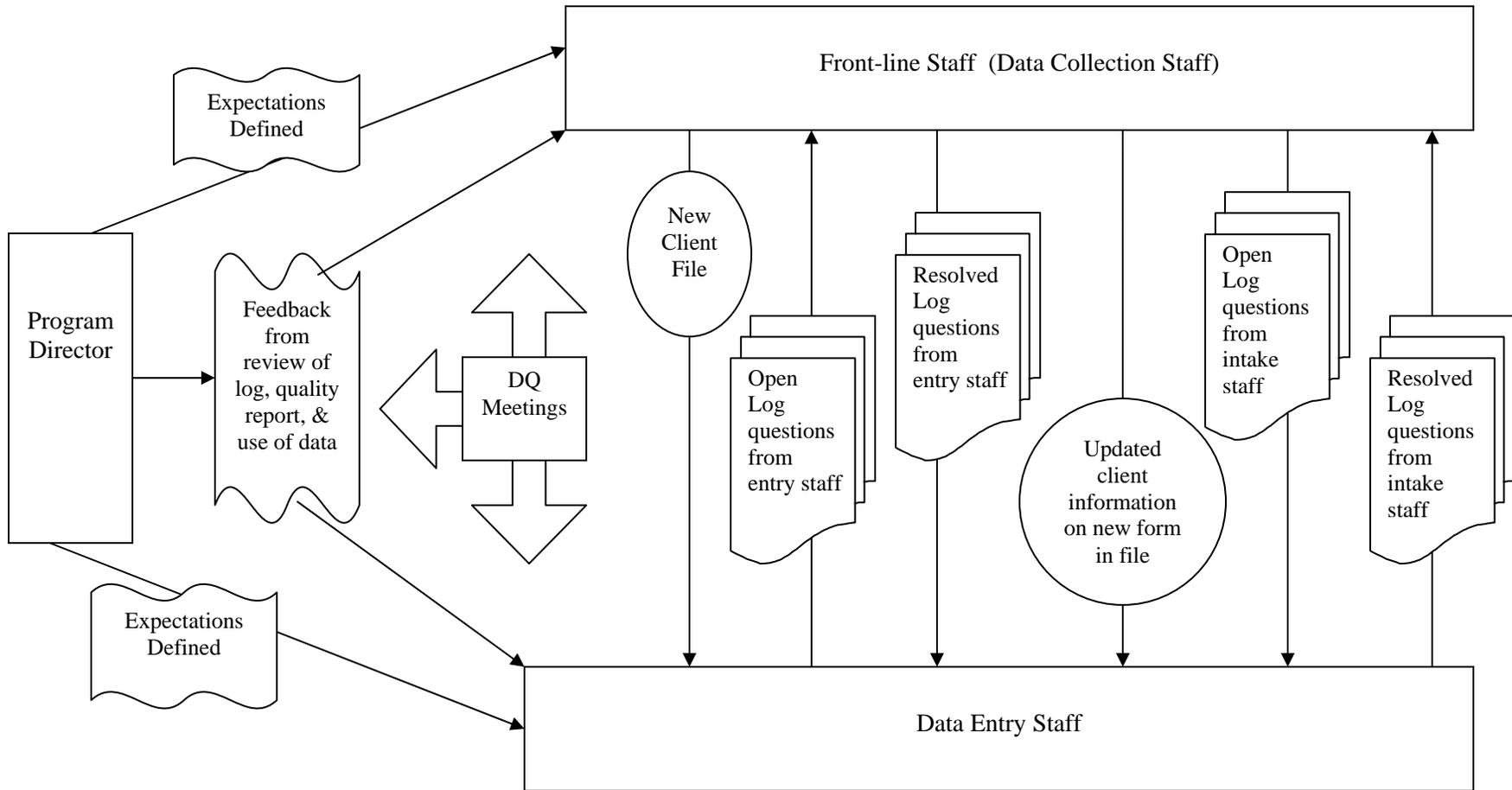
Standard	Benchmark	Questions to Consider	Possible answers**	Decision/Policy
		How do we measure consistency of Q&A's	Random call to interviewers quarterly and ask how questions are worded, 90 percent should be on target	
	Common knowledge of what fields to answer	How do we ensure common knowledge of required fields	Communication! User group meetings every month, constant reminder of essential fields; Screen design (e.g. asterisks, color-coding)	
		How do we measure consistency of knowledge?	Common training forms that list fields; Monthly check of required fields in system – 95% of records have complete minimal fields	
Monitoring	All benchmarks are monitored regularly	Who will monitor compliance with policy?	Supervisor. HMIS Project Manager. Self-monitoring with signed report.	
		How often will monitoring take place?	Once a week. Once a month.	
Incentives	All policies should include incentives for compliance.	Should there be positive incentives? Repercussions for non-compliance? Both?	Funding / job performance ties (Either positive or negative)? Tied to software access. Data entry staff of the month program.	
Agreement	All parties should have an explicit mechanism affirming knowledge of and agreement to quality plan.	How can we ensure that all stakeholders agree to plan?	A written contractual arrangement. Built in consent screen in software. Part of job description.	

** These are suggestions for your CoC, and/or state, region to consider for consistency. But your HMIS implementation should have its own answers to these questions to assure Data Quality and usefulness of data.

Appendix 4: Sample Data Quality Log

Client ID	Staff	Date	Field	Issue	Staff Resolving	Date of resolution	Description of Resolution
56789	JP	1/15/05	Middle Name	Illegible	LGG	1/16/05	JAMES
58943	BF	1/18/05	Disability	What does “SA” stand for?	HS	1/19/05	“Substance Abuse”
574384	JR	1/18/05	Program Entry	Field missing	HS	1/19/05	Should be 1/18
577864	JR	1/18/05	PIN	Client said was here before, but can’t locate in HMIS.	LGG	1/19/05	Now using maiden name, previously served under married name. (added to form).
544753	JP	1/20/05	Race	Is client really Native Hawaiian? (Check marked between boxes).	LGG	1/21/05	Oops. Should be “White”
567789	LGG	1/25/05	??	Couldn’t find client in HMIS using SSN	JP	1/26/05	Typo in both SSN and last name, fixed.
577455	BF	1/31/05	Entry/Exit	Exit earlier than entry. Should these be reversed?			
566674	JP	2/2/05	Ethnicity	Both yes and no checked. Pleas advise			
577735	BF	2/3/05	Household	No children forms are attached.			

Appendix 5: Program Level Data Process and Feedback Loop



Appendix 6: Director's Memo

TO: All Staff
FROM: Max Smedley, Program Director
DATE: 1/1/05
RE: HMIS Data Processes and Quality Procedures

In order for us to benefit from entering data into the HMIS, it is important that we make sure that we are entering all information accurately and completely. This information is critical for us to understand the people we serve and to make our case to funders.

To ensure quality data, our agency has adopted the following policy on data quality:

- Information (with particular emphasis on the minimally required fields) should be entered in the HMIS for all clients, unless the client refuses.
- Overall for the agency, no minimally required data field should have missing information for more than 5% of records.
- Data should be entered into the databases within 48 hours of every client intake or service encounter.
- All active clients should be reviewed by the 5th of every month to ensure their data is still timely.
- Data intake staff should randomly spot check at least 5 of their clients every other week in the HMIS against the paper to ensure that information was entered accurately.
- All questions about the data should be entered in the data quality log. After entering a new question, the staff member to whom the question is addressed should be alerted by email. Responses to queries in the quality log should be made within 2 business days. The data quality log will be reviewed at each weekly case manager meeting.
- We will use client reports generated from the HMIS, rather than paper files when reviewing case files in weekly case manager meetings.
- Data quality will be added as an agenda item in our weekly staff meetings.
- Data quality goals will be added to the performance review criteria for all relevant staff.

The following processes will ensure that data are not missed .

- **New** client forms should be placed in the data entry inbox. After data entry, data entry staff should file the client form,
- Information about **returning** clients or **updated** information about current clients should be recorded on the **blue** update forms, these should be clipped to the top of the client file and the whole file should be placed in the data entry inbox.
- Client **exit** information should be entered by the client case manager.

All staff HMIS users will be expected to implement these policies, as relevant to your position. If you have any questions or concerns about these procedures, please contact me.

Appendix 7: Possible Validations of HMIS Universal Data Elements

Name

- First and Last name not same
- Suffix properly formatted
- No numerals in name fields
- Suffixes not in last name field
- First name is not “Husband,” “Wife,” “Man,” “Woman,” “Boy,” “Girl,” “Child”, “Baby,” “Baby Girl,” “Baby Boy” or similar

Social Security Number/Quality Code

- SSN has all numbers and no dashes
- 9 digits when quality code indicates complete
- Less than 9 digits when code indicates partial
- All digits not same (333333333); all numbers not sequential (123456789)

Date of Birth

- Earlier than current date
- Earlier than program entry date
- Later than 90 years from present
- Not minor in adult shelter/Adult in youth shelter

Ethnicity/Race

- Primary and secondary race not the same

Gender

- Men not pregnant
- No Male in woman’s shelter/Woman in men’s shelter

Veteran Status

- Client under 18 not veteran
- All veterans in veteran shelter
- Those receiving veteran’s pension marked as veteran

Disabling Condition

- Those receiving SSDI for themselves are marked as having a disability
- Those indicating substance abuse, mental health, physical disability, developmental disability, HIV/AIDS marked as having disability

Residence prior to program entry / How Long At Place

- Self-report not contradicted by other HMIS data

Zip Code of Last Permanent Address/Quality Code

- Zip code complete if quality code marked a complete
- Zip code five or nine characters
- Zip code valid (If list of zips available)
- Zip code has only numbers

Program Entry Date/ Program Exit Date

- All clients have a program entry date.
- Program Entry Date later than Birth Date
- Program Entry Date prior to Exit Date.
- Entry and exit date not the same in residential shelter

- Length of program enrollment outliers are reasonable considering program type

Household ID

- Single person in family shelter
- Family in individual shelter

Appendix 8: Sample Data Quality Reports

Aggregate Implementation-Level Report

Source: State of Wisconsin HMIS

This system-wide report calculates total new records and duplicate records in the system. It also looks at the percentage of records that are null for particular fields, based on the total active records. The report is run monthly so systemic progress or fluctuation in data quality can be observed.

Rpt		March	April
Run 01	Ttl # Client Records Active	93896	97838
	Ttl Count -Unduplicated	90053	93501
	Number of duplicates	3380	3804
	Pct of total	3.6	3.9
	Ttl Nmbr Added this Mo.	2587	3942
NULLS			
Run 02	DOB		
	Distinct count with data	84360	87317
	Unduplicated less Distinct	5693	6184
	Pct of Unduplicated - Null	6.3%	6.6%
Run 03	Gender		
	Distinct count with data	87492	90570
	Unduplicated less Distinct	2561	2931
	Pct of Unduplicated - Null	2.8%	3.1%
Run 04	Race		
	Distinct count with data	73950	76700
	Unduplicated less Distinct	16103	16801
	Pct of Unduplicated - Null	17.9%	18.0%

Single Score Per Program Report

Source: Adapted from State of Wisconsin HMIS.

This simplified format creates a single data completeness score per program. The score is calculated by looking at all the clients served in the program and a particular number of fields. If the report tracks data entry on 11 fields, and the program served 50 clients, then the program would receive a score of 100% if data were entered for 550 fields.

Program Name	Score
Program A	100%
Program B	93%
Program C	95%
Program D	86%
Program E	43%
Program F	89%
Program G	77%
Program H	62%
Program I	100%
Program J	80%

Single Program Report Including Additional Guidelines

Source: Community Shelter Board, Columbus, Ohio.

Agency: Homeless Service Provider

Program Name: Family Emergency Shelter

Site Administrator:

Time Period: 07/01/2004-12/31/2004

Date of Review: 03/04/05

		Within Guidelines		Note
		Yes	No	
Total Number of Clients in report	227	X		
Total Head of Household	67	X		
Total Number of Adults in Families	87	X		
Total Number of Children in Families	139	X		
Number of Non HH Family Members	160	X		
Number of Adults not in Families	1		X	
Shelter Units/Daily Occupancy	23 (4296/184)		X	{20} 115%
Total Household Exits	46	X		

Required Data Elements	Completion		Within Guidelines		Not Currently Measured
	#	%	Yes	No	
Client Name					X
Client SSN					X
Date of Birth	67	100%	X		
Gender	67	100%	X		
Race	67	100%	X		
Hispanic/Latino	221	97%	X		
Household Relationship					X
Family Status	67	100%	X		
Family Member Birth Dates	160	100%	X		
Family Member Race	156	98%	X		
Family Member Gender	159	99%	X		
Head of Household	67	100%	X		
Entry Date					X
Veteran Status	66	75%		X	
Education Level	82	93%		X	
Disabled or Handicapped	61	69%		X	
Employment Status	78	89%		X	
Homeless Status					X
Type of Previous Housing	64	96%	X		
Primary Reason for Crisis	65	97%	X		
Secondary Reason for Crisis	64	96%	X		
Previous Zip Code					X

Enhancing HMIS Data Quality

Previous General Area					X
Financial Resources at Intake	42	100%	X		
Income amount at Intake	42	100%	X		
Income Source at Intake	42	100%	X		
Exit Date	46	100%	X		
Reason for Leaving	43	93%	X		
Destination	40	87%		X	
General Area Exit	45	98%	X		
Income Amount at Exit					X
Income Source at Exit					X

Aggregate Reports by Program for HMIS Project Manager

Source: CSPTech Program, Massachusetts

					PERCENTAGE										
Provider	HMIS Staff	#Total	#New In Year	#New In Quarter	First Name	Last Name	SSN	DOB	Gender	Ethnicity	Race	Prior Living	Vet	Disability	Last Perm Zip
Provider A	Joe	1030	354	111	100	100	74	99	99	96	99	96	95	54	62
Provider B	Mary	26	15	10	100	100	100	100	100	100	100	93	93	95	
Provider C	Mary	1	1		100	100		100	100	100	100				
Provider D	Joe	417	135	26	100	100	100	100	100	100	100	100	100	100	100
Provider E	Mary	915	231	3	100	100	91	100	100	94	100	37	99	84	81

Single Program Data Quality Report Sent to Program Directors

Source: CSPTech Program, Massachusetts

HIMS Program-Level Data Quality Report

This report shows the general status of data entry for your program in the HMIS. The report shows how many records had been entered for the year and reporting period and the data quality for these records based on those entered during the year. Please note that records can only be counted for a particular program if the person entering data is attached to that program.

I. What and Where the information is been monitored

We mainly monitor the basic demographic information for each client. All of the monitored fields are located on Required Data Elements Screen, except prior living and reasons for homelessness. These are located on the Residential Assessment Screen. Some fields such as gender, date of birth, etc., can be filled on other assessment screens as well, since all of the data is dynamic and will automatically fill in other sections.

II. Report for [**PROGRAM NAME**] - Quarter 4, 2004

Description of the table:

Total In System -- unduplicated count on total clients that have been entered into system

New In Year -- Unduplicated count on clients that had been added by provider yearly

New In Quarter -- Unduplicated count on clients that had been added by provider yearly/quarterly

All the percentage fields -- percentage of corresponding fields that were filled in for clients added in current year.

Field	Quality
Total in System	1030
New In Year	354
New In Quarter	111
First name	100%
Last name	100%

Enhancing HMIS Data Quality

Date of Birth	99%
Gender	99%
Ethnicity	96%
Race	99%
Veteran	95%
Where Stayed Last Night	96%
How Long at Last Place	96%
City of Last Permanent Address	72%
State of Last Permanent Address	69%
Zip of Last Permanent Address	53%
Disability	46%
Marital Status	97%
Education	96%
Current Employed?	96%
Continuously homeless?	96%
Times homeless in last 3 years	94%
Receiving SSI/SSDI	95%
Medical Insurance	0%
Has Income?	94%
Prior Living	39%
Primary Homeless Reason	0%

Appendix 9: Documents from Michigan State HMIS

Source: Michigan Coalition Against Homelessness

Excerpts from Policies and Procedures

- 4). Data Quality** PPI data will be accurate, complete, timely, and relevant.
- a. All PPI collected will be relevant to the purposes for which it is to be used.
 - b. Identifiers will be removed from data that is not in current use after 7 years (from date of creation or last edit) unless other requirements mandate longer retention.
 - c. Data will be entered in a consistent manner by authorized users.
 - d. Data will be entered in as close to real-time data entry as possible.
 - e. Measures will be developed to monitor data for accuracy and completeness and for the correction of errors.
 - i. The agency runs reports and queries monthly to help identify incomplete or inaccurate information.
 - ii. The agency monitors the correction of incomplete or inaccurate information.
 - iii. By the 15th of the following month all monitoring reports will reflect corrected data.
- 6) Record Access and Correction** Provisions will be maintained for the access to and corrections of PPI records.
- a. Clients will be allowed to review their MSHMIS record within 5 working days of a request to do so.
 - b. During a client review of their record, an agency staff person must be available to explain any entries the client does not understand.
 - c. The client may request to have their record corrected so that information is up-to-date and accurate to ensure fairness in its use.
 - d. When a correction is requested by a client, the request will be documented and the staff make a corrective entry if the request is valid.
 - e. A client may be denied access to their personal information for the following reasons:
 - i. Information is compiled in reasonable anticipation of litigation or comparable proceedings;
 - ii. Information about another individual other than the agency staff would be disclosed,
 - iii. Information was obtained under a promise of confidentiality other than a promise from this provider and disclosure would reveal the source of the information
 - iv. Information, the disclosure of which would be reasonably likely to endanger the life or physical safety of any individual.
 - f. A client may be denied access to their personal information in the case of repeated or harassing requests for access or correction. However, if denied, documentation

will be provided regarding the request and reason for denial to the individual and be made a part of the client's record.

- g. A grievance process may be initiated if a client feels that their confidentiality rights have been violated, if access has been denied to their personal records, or if they have been put at personal risk, or harmed.
- h. Any client grievances relative to HMIS will be processed/resolved according to agency grievance policy.
- i. A copy of any client grievances relative to HMIS data or other privacy/confidentiality issues and agency response are forwarded to MiCAH.

Quality Assurance Policy and Procedure (2004)

POLICY AND PROCEDURE

DATE: September 1, 2004

SUBJECT: MSHMIS Quality Assurance and Data Quality

REASONS FOR POLICY:

- To insure that data is accurate
- To identify and problem solve barriers

STATEMENT OF POLICY: MSHMIS will maintain and on-going process of Quality improvement. This process will be built around routine user meetings that occur at multiple levels of the implementation and routine measurement of data quality and outcomes related to mission critical processes.

2) User Meetings:

a) Purposes:

- A. Opportunity for Benchmarking between participants:
- B. Review core processes and related measures.
 - (1) Identify issues and share solutions.
 - (2) Identify those issues where additional help is needed.
 - (3) Incorporate process and outcome measures. (For example MSHMIS requires coverage rates types of reports generated at the agency level)
- C. Support transparency.
- D. Share successes.
- E. Review aggregated data
- F. Formalize communication to database administrators.
- G. Provides routine consumer satisfaction input.

b) Types and frequencies:

- A. Agency User Meetings - quarterly.
- B. Agency Administrator / CoC User Meetings – monthly.
- C. System Administrator User Meetings – quarterly.
- D. Specialty Provider Meetings:
 - (1) Domestic Violence
 - (2) Runaway Youth
 - (3) Housing Specialists
 - (4) Mental Health
 - (5) Homeless Education Providers

- c) Meeting Requirements
 - A. Minimum attendance levels (= all those with user licenses + leadership)
 - B. Structured Meeting Agenda reviewing core processes:
 - (1) Coverage – Are all the clients being entered? What % of the homeless are in the System?
 - (2) Client Refusals
 - (3) Data Quality – null data fields, # of data corrections
 - (4) Interview issues
 - (5) Definition questions.
 - (6) Training needs
 - (7) Privacy and Security issues
 - (8) Reports: review Agency aggregated data
 - C. Structured Minutes with copy sent to database administrators to monitor user meeting compliance with the MSHMIS contract.
 - d) Additional Processes related to System Admin and MSHMIS Meetings
 - A. System Access / Licenses
 - B. System Performance
 - C. Routine Support / Help Desk
 - D. Contract Compliance (Provider and Vendor)
- 3) Sources of Data:
- a) Consumer input from structured minutes
 - A. Qualitative and Quantitative
 - b) Measures that result from routine queries within the database.
 - c) Short Term measures to assess problem resolution.
 - d) Consumer Satisfaction Surveys
- 4) Improvement Cycle:
- a) Assess variation /one time variation due to unique conditions or sustained issue.
 - b) Prioritize problem solving.
 - c) Define a plan for change.
 - d) Test the success of your changes.

5) Measurement Grid:

Process	Measure (control chart / sentinel event)
Coverage	<ul style="list-style-type: none"> • Agency -% of planned entry completed -monthly • CoC-% of housing chart covered – annual. • State -% of CoCs contributing - annual.
Training	<ul style="list-style-type: none"> • # of individuals trained. • Follow-up Satisfaction Surveys (Minutes). • # of records identified in data integrity reports.
Privacy	<ul style="list-style-type: none"> • Grievances or breaches. • Privacy Issues identified on Minutes. • % of clients entered as anonymous. • % or providers with profile closed.
Performance	<ul style="list-style-type: none"> • Avg System response time on “saves” and “report generation” sampling weekly at different times. • # of system bugs reported to QI at Vendor.
Help Desk	<ul style="list-style-type: none"> • % of help desk inquiries answered during call, within 1 day, within 3 days, & within 1 week. • % of unresolved or follow-up Help Desk requests. • Consumer Satisfaction – minutes.
Access	<ul style="list-style-type: none"> • Audit of agency training log forms against users in the System • # of license/user variations.

6) Date Quality:

a) Data must pass “Fitness for Use” Tests

A. Completeness

- (1) Information is entered on all consumers.
- (2) Information on the consumer is complete.

B. Accuracy

- (1) Data reflects reality.
- (2) Data is entered correctly.
- (3) Data has face validity – reflects what we know.

C. Consistency

- (1) Performance information is consistent across time.
- 7) Common Errors:
- a) Systematic Errors/ Issues with Training:
 - A. Entering “no” when you mean “yes.”
 - B. Definition drift.
 - C. Entering text without using drop down.
 - D. Failing to enter information on some consumers.
 - b) Random Errors/Sloppy Entry/Workflow
 - A. Date Errors (DOB is 4/15/52, entered 4/15/04)
 - B. Transposing numbers
 - C. Spelling errors (Lauren vs Loren)
 - D. Accidentally selecting the wrong response from a drop down.
- 8) Factors impacting quality:
- a) Prioritized Process in the Organization?
 - A. Are staff given the time to participate in training and to complete entry?
 - B. Is the environment arranged to support entry?
 - C. Is the process owner within the agency respected?
 - D. Is the data used?**
- 9) Consumer Procedures for ensuring Quality:
- a) Standardized collection instruments.
 - b) Creating an environment conducive to data collection and entry.
 - c) Event triggers for data collection and entry – clearly defined work flow.
 - d) Guidance for special populations.
 - e) **Must run reports monthly!**
- 10) HMIS Procedures for ensuring quality:
- a) HMIS staff monthly reviews reports for completeness, accuracy and consistency.
 - b) Clear protocols for correcting data.
 - A. Agency signs off on reports monthly.
 - B. Errors systematically result in corrective action.
 - C. Procedures for correcting are defined.
 - c) Software has error checking functions (out of range, missing values, incongruous data).
 - d) Staff look at data reliability and validity issues prior to publishing reports. Collecting agencies will know which questions result in data that simply is not stable. Do the findings make sense? Must be knowledgeable about local services to recognize systematic data errors.
 - e) **Using the data.**
- 11) Measures to Monitor Quality:
- a) Queries:
 - A. Null DOB and gender fields.
 - B. Rate of infants under the age of 1.
 - C. Gender by family relationship.
 - D. Homeless by “extent of homelessness.”
 - E. HUD Assessment by entries & exits.
 - F. Age by family relationship.
 - G. Number of users, CoCs, and records on the live site.
 - H. Null exit dates related to short term services.
 - I. Ambiguous data in reports

Appendix 10: Sample Statement of Limitations, Coverage Rates, and Parameters

Source: CSPTech, Massachusetts

These data provide information about individuals served in Massachusetts non-specialized shelter programs; they do not necessarily reflect the characteristics of the homeless population overall. These data do not capture information on families, people who are in doubled-up living situations, and others who are homeless but do not come into contact with the service system.

In particular, individuals utilizing the following types of services are usually not represented in these data: street outreach programs, domestic violence shelters, substance abuse treatment and detoxification programs, hotels, programs serving persons living with AIDS, healthcare programs, (transitional housing programs), and permanent housing programs. Individuals who are deemed ineligible for service are also not included in these data.

It should be noted that the homeless individuals who are interviewed as part of this project respond to specific questions with varying levels of depth. As can be seen in the graphs in the report, information on demographic characteristics is based upon a much higher response rate than for city of prior residence, income sources, and other data fields.

In addition, the numbers of valid responses across various questions are often quite different and vary from year to year. For example, the number of respondents in 2001 for prior living is almost double the number of respondents in 2000 and 1999. The number of respondents increased by nearly 50% in 2002, and almost doubled between 2002 and 2003 thanks to increased efforts in data collection. While these cross-year comparisons provide noteworthy information, the variations in response rate should be taken into account when making generalizations about the data. The results, can, however, still provide some indication of the differences among homeless populations, shelter guests in particular, across the five years.

Based upon policies developed by the project's Steering Committee, aggregate data must meet a minimum threshold criterion before they can be released: data must represent at least 60% of those persons served by the emergency shelter system in a region.²⁰ Based on a calculation of client records contained in the database versus shelter system capacity for a particular period, the data are deemed eligible for release.

This Massachusetts CSPTech report represents information on individuals who utilized the state emergency shelter system in 2003. The data presented in this report represent 65% of individuals served at nonspecialized Massachusetts homeless shelters between January 1, 2003 and December 31, 2003.

²⁰ There will be some cases where a "slice" of aggregate data does not meet the 60% test; however the available data are needed to support an important policy debate about an identified trend. Slices of data not meeting the 60% threshold require a judgment call; in those cases a three person Access To Data executive committee (a member representing homeless families, another representing homeless individuals, and another representing funders of the system in Massachusetts) will be consulted. This committee then decides whether data not meeting the 60% test will be publicly released.

These coverage rates are determined by calculating the total number of beds in the shelter system, and multiplying that figure by the average annual turnover in those beds, thus estimating the proportion of total persons served by the shelter system represented in the data. For example, if the individual shelter system has 4,000 beds across all of the nonspecialized emergency shelter programs, using a turnover rate of 5, the shelter system would serve 20,000 persons over the course of the year. If there were 12,500 individual records for the year, coverage would be 12,500/20,000, or 63 percent. For individuals, the Steering Committee agreed (as specified in the Three Year Workplan) on a turnover rate of 5, the average of the actual turnover reported by Dennis Culhane in Philadelphia and New York City in 1994. In 2000 this rate was increased to 6.5, and in 2003 the Massachusetts rate was again increased to 12, based upon actual data from the programs showing that individual stays are shorter in Massachusetts than in the other two cities.

$$\begin{array}{l}
 \text{Total MA Beds} \times \text{Turnover Rate} = \text{Total MA Persons Served} \\
 \text{Total CSP Records} / \text{Served} = \text{Coverage Rate}
 \end{array}$$