2012

Data Integrity Team,
working team of the
Homeless Management
Information System and
Data Technical Assistance
Workgroup

[PRIORITY RECOMMENDATIONS FOR WORKPLAN DEVELOPMENT]

This document represents the work 5 different National HMIS Technical Assistance Providers meeting over a period of 2 months. Individual members were identified by firms based on their specific experience related to HMIS operations. Where possible the team has identified both the timeline for product development and the specific audience for the product. The Data Integrity Team strongly recommends that it be assigned the role of product reviewer to ensure fidelity of the end product with the original intent.

1. **Automated Client Exit in HMIS**

   Key HUD reports, such as the Annual Performance Report (APR) and the Annual Homeless Assessment Report (AHAR), rely on the ability of homeless assistance programs to report client lengths of stay. Additionally, with the emergence of community-level performance measurement through the implementation of the HEARTH Act, the ability of all programs within a CoC to accurately measure length of stay will prove even more critical. Even so, it is often very difficult for certain programs (specifically emergency shelters) to provide correct length of stay data given the transient nature of their clients.

   This topic will develop HUD guidance for HMIS solution providers on the use of automatic client exits (i.e. auto exits) andTA resources and community best practices on the development of auto exit standards, to help both programs and communities better understand how the use of auto exits mightimprove their HMIS dataand when it would be inappropriate to generate auto exits.

   <u>Proposed Products</u>

   a. White Paper (2 Months Development)
      Documentation of known best practices around auto exits, as well as sample standards for auto exits;
   b. HMIS Software Provider Guidance (3 Months Development)
      Written HUD guidance for solution providers on recommendations for auto exiting procedures.
      i. Specifications for recommended alerts in an HMIS system document
      ii. Specifications for auto exiting document
   c. Webinar (2 Months Development)
      30 minute webinar for HMIS System Administrators on how to implement autoexits into their local HMIS Implementation, should include discussion of topics covered in White Paper.

2. **HMIS Automated Data Quality Tools**

   This product would discuss and analyze the role of automated tools in the improvement of data quality. These tools include automated error alerts to notify HMIS users of missing or potentially erroneous data, as well as required data entry, which prevents users from saving forms that are missing responses to one or more critical data elements. Both have been implementedin a variety of HMIS products. Critical examination of the effect of these mechanisms on the data produced by HMIS will allow Continua of Care to implement these tools more effectively and identify the practices with the most impact.

   <u>Proposed Products</u>

   a. Software Provider Guidance
      White Paper identifying the alerts or required data entry practices that produces the greatest increase in data quality. Work with a handful of Continua to examine the impact of these mechanisms on data quality.

b. Webinar/Training
   Webinar discussing the results of the white paper combined with a presentation on what areas Continua may prioritize to produce the most impact. Audience for webinar includes HMIS administrators and technically savvy audiences.

3. **De-duplication Procedures**

The HEARTH Act calls upon Continuums to measure the effect of their interventions on average length of stay in homelessness, recidivism (returns to homelessness after placement in housing), and the incidence of new homelessness. All of these statistics depend upon the ability to recognize duplicate records in HMIS that identify the same client and to manage them such that it is possible not only to provide an unduplicated count of people experiencing homelessness, but also to track individuals' movement through a Continuum's programs. To the extent that HMIS implementations across the country are using different approaches to de-duplicate, it becomes impossible to accurately compare the performance of one CoC to the performance of another. A robust de-duplication process will be more likely to identify returning or re-entering clients as such, and, therefore, will show more accurate lengths of stay and recidivism rate than a de-duplication process that fails to recognize that the clients were in the system before.

De-duplication is the process of sorting through identifiers to determine whether the two or more identifiers correspond to the same individual. Any de-duplication process should be able to pick out matches involving identical identifiers. Where two (sets of) identifiers are not exactly the same, the robustness of the de-duplication process is a measure of its ability to maximize the percentage of instances in which actual matches are correctly identified, and to minimize the percentage of instances in which two distinct individuals are presumed to be the same person.

Different rules for duplication and identifier matching may be appropriate for different situations:

(a) during data entry, when a user enters identifiers which at least partially match the identifiers of a client already in the system, it may be appropriate to use an algorithm which has a lower threshold for determining a match, so that any possible duplicate can be identified;
(b) during data checking, when an administrator searches the data base for duplicates, it may be beneficial for the software to establish a lower threshold for determining a match, so that any possible duplicate can be identified; and
(c) during the running of a report, it may be more appropriate for the software to maintain a higher standard for identifying matches
(d) during uploading of data which is to be integrated with existing data, it is appropriate for the software to use the highest standard for determining a match, so as to avoid integrating the records of different clients.

<u>Proposed Products</u>

a. Software Provider Guidance
   Clear guidance that includes the role of human decision making to software providers to implement specific mechanisms for de-duplication preferably through the upcoming Notice. In event that this is not possible then use of White Paper to

define one, or a set of, highly recommended mechanisms for de-duplication. The software should be transparent about the basis for identifying matches between identities; so that it becomes possible to compare two different datasets in terms of how likely they are to generate false positives vs. false negatives. Work should build on paper written by Brian Sokol and Oscar Gutierrez in 2005.

b. Webinar/Training
Discussion of theimportance of de-duplication both during data entry and during report generation. This training would be targeted to multiple audiences.

4. **Script-based Data Alteration**

CoCs struggle with the creation of quality data on a regular basis. A simple error in training or procedures can cause hundreds or thousands of records to be created improperly. The constant struggle of frontline staff to prioritize data entry leaves many datasets with missing data related to exit information. As CoCs grow more sophisticated, and HUD penalizes CoCs through the NOFA more significantly for lack of HMIS reporting, the use of scripts and other software based tools to populate records may become a substantial barrier to the collection of accurate data. Already, with the best of intentions, some HMIS Administrators are populating exits dates for entire CoCs based on the CoCs' average length of stay. The implications for this practice are not fully understood however continued expansion of this practice couldundermine conclusions drawn by policy makers and produce poor outcomes.

<u>Proposed Products</u>

a. White Paper (4 months)
Describe current practices within CoCs using software based tools to assist with data entry. Include the instructions provided to software providers by the HMIS administrators and if possible, the impact on local reports.

b. HMIS Admin Guidance/Tool (9 months)
Requirements for documenting script-based data alteration in audit logs;

5. **Generation of Unique Identifiers**

Different software use different mechanisms to generate unique identifiers. Software which generates unique identifiers by combining prescribed portions of component identifier fields may hamstring the de-duplication potential of that software, in that matches of such composite identifiers may require that all identifiers between two identities are identical, which would exclude significant numbers of matches in which one or more identifier is missing or entered erroneously.

<u>Proposed Products</u>

a. Software Provider Guidance
Clear guidance via the upcoming Notice, if possible, discouraging the formation of composite identifiers. In event that this is not feasible a White Paper to discuss the impact of use of composite identifiers on the robustness of de-duplication is

advised.Variance in HMIS software practices has an impact on AHAR conclusion and all HUD reporting. This work should build on paper written by Brian Sokol and Oscar Gutierrez in 2005.

6. **Data Merging**

As communities continue to use and expand HMIS over time it becomes essential to combine records that reference the same client and were created erroneously. With large HMIS implementation duplicate client records can burden data entry staff, HMIS report writers, and computing resources. The creation of guidance to HMIS administrators around how to make consistent decisions regarding the merging of conflicting data, more timely answers, or intentional choices around disclosures by clients is critical to the continued growth of HMIS.

Proposed Products

a. White Paper
   A brief overview of the issues combined with the current best practices of communities that have engaged this issue and created comprehensive solutions that account for data maintenance over time.
b. Tools
   Decision matrix providing HMIS Administrators with concrete decision making recommendations and sample exercises to assist in the development of local policy making and database updates.
c. Software provider recommendations

7. **Legacy Guidance**

With the impending revision of HUD's HMIS Data Standards, all HMIS solution providers will need to ensure that data entered into their system through the 2004, 2009 or 2010 HMIS Standards is either updated to match the format of the revised standards or otherwiseable to be reported out in the structure of the new elements. Failure to map data from the old to new standards can result in missing data on HUD or non-HUD reports, when in fact the data was entered into HMIS. For example, under the 2010 HMIS Standards "Rental by client, VASH subsidy" was an allowable response to the Destination data element (4.10). However, under the revised HMIS Data Standards, no such response will exist. Therefore, HMIS solution providers will need to program into their HMIS a method for data entered as "Rental by client, VASH subsidy," to appear as the appropriate category in the revised standards.

In addition, alterations to HUD formulas, reporting parameters, and report formats may have significant impact on fluctuations of local reports. HUD currently offers no guidance on how CoCs should address the validity of prior HMIS based reports that deliver different outputs than when the report was originally created.

Proposed Products

a. Software Provider Tool (3 months)
   Crosswalks that 'translate' old response categories to 'new' for data alteration and/or reporting purposes;
b. HMIS User Tool/Training (3 months)
   Webinar, handbook, or other training to explain what changes HMIS users may see and why); and
c. Software Provider Tool (6 months)
   Data mapping procedure guidance that standardizes the methodology HUD will use in the generation, implementation, and maintenance of prior reports.

8. **Internal HUD/TA Report Release Checklist**

Supporting CoCs with the creation of valid, useful reports that provide insight into program impact, CoC progress, and assist with serving clients must be prioritized throughout the creation and distribution of HUD reporting requirements. CoCs and software providers often require months of advance notice and planning to implement changes that are mistakenly perceived to be minor. The creation and adoption of a transparent checklist would assist in rebuilding confidence and trust among HUD clients and significantly simplify the release schedule for reports as all stakeholders expect and know the entire process and are clear on what role they play.

Proposed Products

a. Checklist (4 Months)
   Creation of checklist that combines contributions from software providers, HUD report writers, and HMIS TA members to ensure broad participation. Checklist should provide clear benchmarks with dependable intervals between tasks. While final report release is subject to significant interference from multiple variables often the creation of the report is not.

9. **Report Data Quality Error Reduction**

The creation and use of quality data is one of the highest priorities of HMIS technical assistance. Without this foundation the evaluations, outcomes, outputs, and ultimately decisions made may be flawed. HMIS TA must continue to focus on the constant struggle to produce the cleanest reports possible given the resources and constraints. Critical to CoCs and HUD is the creation of Annual Performance Reports (APRs) for use in program evaluation. Several tools can be made available to CoCs, software providers, and HMIS administrators to assist in the early identification of potential data quality errors. Tools in the software, training techniques for users, and targeted reports will all support the CoC to more effectively use HMIS.

Software providers could set up the real-time alerts, program diagnostic reports, and raw data APRs, and would do or facilitate the programming of queries which could be pasted

into pre-programmed spread sheets that could diagnose possible errors, inconsistencies, etc. Alternatively, the pre-programmed queries could be available on the HUD website, and could be used in conjunction with any HUD-compliant HMIS software, because part of HUD compliance would be the ability to generation the appropriate data sets (e.g., via queries).

Proposed Products

a. Tools
   Creation of standardized report logic and spreadsheets for use by HMIS administrators in the identification of likely data errors.

   Alternately, it software vendors can generate defined-field spread sheets containing the data upon which an APR (or AHAR) will be based, that spread sheet could be pasted into a standard diagnostic tool to assess the prevalence of errors or actual or potential inconsistencies.
b. Tools
   Creation and distribution of a specific list of frequent data entry errors for use by software providers and HMIS administrators in developing controls.