Accurate, Complete, Timely HIV Surveillance Data for Public Health Action

Technical Guidance for HIV Surveillance Programs

Data Management

HIV Surveillance Branch
Division of HIV Prevention
National Center for HIV, Viral Hepatitis, STD, and TB Prevention
Centers for Disease Control and Prevention
Atlanta, Georgia

The HIV Surveillance technical resources are being modified to comply with Executive Orders and the OMB Paperwork Reduction Act.

Please direct all questions and requests regarding resources on the HSB SharePoint to HIVSurveillance@cdc.gov

Acknowledgments

The *Technical Guidance for HIV Surveillance Programs* is a series of Technical Guidance files that are part of a portfolio of resources to guide HIV surveillance programs at health departments in U.S. states, cities, and territories on the implementation of HIV surveillance systems in accordance with state, local, and territorial laws, regulations, and practices.

These files are living documents and the updates include adaptions and adjustments from previous iterations. We acknowledge previous contributors from the Centers for Disease Control and Prevention (CDC), other federal agencies, academic partners, and state, local, and territorial health departments.

The updates to these files were prepared by CDC, Division of HIV Prevention, in collaboration with the Council of State and Territorial Epidemiologists (CSTE), with contributions from state, local, and territorial HIV surveillance programs.

Contents

Introduction 4
Structural Requirements 5
Surveillance Standards 6
Process Standards 6
Data Management Tasks 7
Data Entry 7
Paper documents 7
Electronic documents 7
Record Linkage 7
Record Management 7
Multiple documents per case 7
Editing documents 8
Processing follow-ups, corrections, and deletions 8
Follow-up 8
Data corrections 8
Case deletions and merging 9
Data Transfer 9
Transferring data to CDC 9
Data Storage 9
Storing data for the local program 9
Additional Documentation 10
Data Management Process Modifications 10
Data Governance Documentation 10
Database Management Documentation 11
Data Security Documentation 11
Data Quality Documentation 11
Reference Documents 11
Other Documentation 12
Records Management Documentation 12
Creating Data Management Reports 12
References 13
Appendix A 14
HIV Surveillance Primary Identifiers (e.g., STATENO or CITYNO) Assignment 14
Appendix B 15
Managing and Maintaining Facility and Provider Data in a Document-based Surveillance Information System 15
Facility/Health Care Provider Creation in the Surveillance Information System 16
Facility Naming Convention 17
Facility Type Assignment 17
Facility Consolidation 18

Introduction

The goals of HIV surveillance are to describe the burden and epidemiology of HIV, monitor HIV trends, identify HIV clusters and outbreaks, and guide public health action at the federal, state, local, and territorial levels.

The *Technical Guidance for HIV Surveillance Programs* is a series of Technical Guidance files that are part of a portfolio of resources to guide HIV surveillance programs at health departments on the implementation of HIV surveillance systems in accordance with state, local, and territorial laws, regulations, and practices. These files are prepared by the Centers for Disease Control and Prevention (CDC) in collaboration with the Council of State and Territorial Epidemiologists (CSTE), with contributions from state, local, and territorial HIV surveillance programs. These files are living documents and the updates include adaptions and adjustments from previous iterations.

State, local, and territorial HIV surveillance programs at health departments are responsible for their HIV surveillance system, which encompasses surveillance activities, reporting sources, surveillance information systems (including the enhanced HIV/AIDS Reporting System [eHARS]), and other supporting tools (like ATra Blackbox, Secure HIV-TRACE). All HIV surveillance systems together contribute to the National HIV Surveillance System (NHSS).

CDC provides technical assistance and support to HIV surveillance programs to ensure that HIV surveillance systems have complete, accurate, and timely data for public health action. The HIV Surveillance Branch (HSB) of the Division of HIV Prevention (DHP), National Center for HIV, Viral Hepatitis, STD, and TB Prevention (NCHHSTP) is responsible for NHSS.

The Technical Guidance files, periodically, will include examples of how eHARS can be used for collecting and managing HIV surveillance data. Note that the same demonstrated concepts should also apply if another surveillance information system is used by the HIV surveillance program.

Technical Guidance file *Data Management* focuses on data management activities for all data sources required for state, local, and territorial HIV surveillance programs without regard for whether the program is integrated with other disease surveillance programs. The activities described in this file include the management of paper and electronic documents, record linkage, and documentation of procedures. These practices are supplemented by the technical manual for eHARS, the eHARS Technical Reference Guide [1].

Considerations when reviewing this file:

Data management is the process of developing policies, procedures, and practices for managing HIV surveillance data and for documenting them on a regular basis. Policies and procedures should include:

- data governance documentation
- database management documentation
- data security and confidentiality documentation
- data quality documentation
- reference documents
- records management document
- other documentation (e.g., creation and retention of analytic data sets)

Proper data management practices can improve quality and accuracy and, at a minimum, safeguard the reproducibility of HIV surveillance data products (e.g., reports, bulletins, fact sheets, slides).

This file will focus on data management activities for all data sources required for state, local, and territorial HIV surveillance programs without regard for whether the program is integrated with other disease surveillance programs.

HIV surveillance programs should ensure that their policies and procedures include the activities described in this Technical Guidance file and relevant public health actions; policies and procedures should be reviewed at least annually and updated as needed. Ensure that staff are trained on the policies and procedures with a focus on changes to procedures or areas needing improvement. Contact the HIV surveillance program's CDC surveillance epidemiologist with any questions or feedback about this Technical Guidance file. If needed, call the HSB main phone number at 404-639-2050 for assistance with identifying the HIV surveillance program's CDC surveillance epidemiologist.

Structural Requirements

Minimum structural requirements for data management by HIV surveillance programs include:

- Capacity to receive and process incoming documents that are submitted by various report sources and of various formats (paper forms, electronic reports, records from other databases).
- Hardware and software that allow for the following activities:
    - searching for cases
    - entry and retention of multiple documents for each case
    - electronic importation of documents
    - data quality control (e.g., data edit checks)
    - report generation
    - linkage with other databases
    - transmission of data
    - backup and restoration of data
- Activities carried out in accordance with CDC's *Data Security and Confidentiality Guidelines* (available at https://www.cdc.gov/program-collaboration-service-integration/php/data-security/index.html) [2].
- Adequately trained staff (initial training and continuing education) in assigned tasks to submit timely (e.g., month end) data transfers to CDC to ensure consistent access to high quality HIV surveillance data. For additional details, see the Data Governance Documentation section.
- Written policies and procedures for each data management activity (i.e., searching for cases, entry and retention of multiple documents for each case, electronic importation of documents, data quality control, report generation, linkage with other databases, transmission of data, and backup and restoration of data). Such documentation can serve as training materials and reference documents when questions regarding processes and procedures arise. All

documentation should be kept up to date to ensure their relevance and usefulness and should be available to all the relevant staff members.

- Primary and backup staff assignments for all data management activities. Having a backup staff member identified to support data transfers to CDC is important. The alternate backup staff should have all the training necessary to independently carry out the data transfer tasks in the absence of the primary staff member. All the necessary access authorization should be obtained and kept up to date all the time.
- Processes for data governance. Data governance encompasses the responsible parties and a set of processes and procedures that ensure the proper management of data throughout the HIV surveillance program (regardless of program integration). The governance documents should describe the program's data management processes and procedures and include roles and responsibilities. When developing data governance plans, the HIV surveillance program should solicit input from stakeholders and public health partners.
- Data manager with thorough understanding of the HIV surveillance system, including the surveillance software and all policies, procedures, and activities that keep the data organized, secure, available, and useful. The data manager should have a solid understanding of all aspects and nuances of the data. Although the person designated as the overall responsible party (ORP) is a public official with sufficient authority to make decisions about statewide HIV surveillance operations and is responsible for the security and confidentiality of the HIV surveillance system [2], this person should not necessarily fulfill the role of the data manager.
- Assigned data management tasks. Managing data may require a team effort. Based on the number of staff members in the HIV surveillance program, one person may be assigned to perform all data management tasks, or the tasks may be assigned multiple members. If multiple assignments are made, it is recommended that one person (i.e., data manager or analyst) be responsible for overseeing and coordinating all data management activities to ensure they are implemented, monitored, evaluated, and when necessary, revised.

Surveillance Standards

Process Standards

- At minimum, conduct an end-of-month transfer of the eHARS data file to CDC by the established deadline (no later than 10am ET 3 business days before the end of each calendar month)
  - All HIV surveillance programs should take necessary efforts to ensure data are complete, accurate and timely before each end-of-month transfer file is submitted to CDC.
  - HIV surveillance programs can perform multiple HIV surveillance data transfers to CDC in a given month.
  - The timeliness of HIV surveillance data is crucial for program planning.
- At least annually, ensure the HIV surveillance program's policies and procedures for data management are documented and updated as needed as part of the overall annual review of policies and procedures described in Technical Guidance file *Overview of HIV Surveillance and Technical Guidance Files*.

Data Management Tasks

Data Entry

Paper documents

Paper documents received at the health department should be marked with the date of receipt. After a visual edit check (proofreading), HIV surveillance program staff should enter the data in the surveillance information system. HIV surveillance programs should document procedures for handling and storing paper and electronic documents, including an acceptable timeframe (e.g., laboratory results are entered within (<) 30 days of specimen collection date) from document receipt to data entry. In addition, HIV surveillance programs should document any program-specific guidance or procedures related to how fields should be completed in eHARS.

Electronic documents

Electronic document collection combines abstracting, coding, data entry, and editing into one process and may be implemented by providing software to the reporting facilities or laptops to health department field staff. HIV surveillance program staff should enter date of receipt of all electronic data.

Importing data from other reporting systems (i.e., ancillary surveillance systems, integrated disease surveillance systems, etc.) or sources (i.e., Electronic Lab Reporting, Electronic Medical Record reports, Electronic Case Reports, etc.) is a critical surveillance function. eHARS is designed to allow for data imports through the use of the All-Document Import function. HIV surveillance programs should routinely (at least weekly) import data reported from laboratories, clinics, hospitals, other public health programs, and other local health departments. Data import templates should be created for each data format, for each document type, and each report source; and they should be reviewed and updated as needed with the release of each new version of the eHARS.

HIV surveillance programs should establish procedures for **verifying keyed data** (see Technical Guidance file *Evaluation and Data Quality*).

eHARS is designed to facilitate case investigations. Documents for potential new cases of HIV infection and those for cases of established HIV infection should be entered, and progress in obtaining the case information should be tracked (see the Processing follow-ups, corrections, and deletions section).

Record Linkage

Record linkage refers to the method of identifying and linking records that correlate with the same entity within one or across several data sources. Record linkage activities are critical for HIV data quality and support efforts to address syndemics, social determinants of health, and health equity. See Technical Guidance file *Record Linkage* for more detailed guidance.

Record Management

Multiple documents per case

Information from multiple documents needs to be consolidated to yield one case record for analysis. If the surveillance information system manages multiple documents per case, most of this consolidation will be

automatic; specifications (hierarchy rules) for this process are outlined in the eHARS Technical Reference Guide, Person View Hierarchy [1]. However, edit checks or other quality control procedures may reveal errors in documents, and the HIV surveillance program's data management staff (e.g., Surveillance Coordinator, Data Manager, or those granted the override authority in the surveillance information system) may want to override the hierarchy rules to ensure that the value selected for a given data field is correct.

Editing documents

Data validations may be automatically applied when information is entered in a surveillance information system or applied to all data on a routine schedule (see Technical Guidance file *Evaluation and Data Quality*). The validations performed in eHARS are described in the eHARS Technical Reference Guide, Person View Edit Checks [1].

For surveillance information systems in which multiple documents can be entered for a case, each document and its information should be retained as originally received. If data validations indicate incorrect information in a particular document, all identified differences should be discussed with each of the original reporting sources. If errors in reporting are detected, changes should be made to reflect the corrected report information. If the original source confirms the data are correct, no changes should be made to the entered document. Data should not be changed solely based on information indicated from other reported documents.

If multiple documents have been entered for a case, rules should be applied to select the most accurate value for a specific item (see the Record Management—Multiple documents per case section). If a value is proven to be incorrect, the option to override that value should be available.

Processing follow-ups, corrections, and deletions

Follow-up

In many instances, follow-up is required to obtain all necessary information on a case or for verification of data. Information obtained from a source during follow-up for missing information may range from one particular data item (e.g., risk factor information) to many data items.

Follow-up information obtained from a new source should be entered on a new case report form (CRF). The CRF should contain the new Document Source Code (e.g., A.01) and the Surveillance Method should be Follow-Up (F). If the follow-up is a revisit to a site or a source has been previously contacted and new data has been collected, the additional information should be entered on a new CRF.

Data corrections

If the information is incorrect (e.g., a data edit revealed an invalid value), but the correct information is obtained by recontacting or revisiting the report source, both paper and electronic records should be updated with the correct information. Changes to original documents should be made only if verification is obtained from the original source.

If information is incorrect on one document but is known to be correct on another document for the same person, staff can ensure that the correct value is used for analysis by reviewing the information on the Person View document in eHARS. If necessary, the value can be corrected by using the Person View override feature [3, Overriding Values]. Data entry errors should be corrected in the existing electronic document.

Case deletions and merging

Staff may need to delete documents or entire cases. For example, a record should be deleted when a person is found not to be HIV infected. When a person has been assigned more than one primary identifier (i.e., STATENO or CITYNO) and a duplicate case has been entered for the person, the data manager can merge the documents in eHARS. After a document or record for a person is marked for deletion or merged with another document, the updated information is sent to CDC and a final deletion is made after feedback from CDC is received that the document/person record deletion or merging has been noted and that the record has been deleted or that the reports have been merged in NHSS. To permanently remove the record from eHARS, the record must be purged [3, Purging Documents].

Data Transfer

Transferring data to CDC

Sending monthly transfers to CDC is important to update NHSS and for creating national HIV surveillance products. HIV surveillance programs should enter all the received documents in eHARS before building the new transfer file. The monthly transfer deadline is three business days before the end of the month and by 10 A.M. ET. All file transfers initiated before the transfer deadline are considered as meeting the process standard.

HIV surveillance programs can conduct multiple data transfers in a given month; however, there should be a transfer closer to the monthly transfer deadline to include all the documents and updated information for a case. The monthly transfers are uploaded to CDC by using Secure Access Management Service (SAMS) Automated Data Transfer (ADT) activity. Once the transfer file is processed, the acknowledgment file should be downloaded from ADT to process in eHARS before building another transfer. Processing the acknowledgment file in eHARS helps reduce the transfer file size by not including unmodified documents that were already transmitted to CDC.

HIV surveillance programs should review each ADT acknowledgment report email to make sure the Person/Case Information totals and the Document-based Information totals are representative of the expected case and document changes made by the program for the transfer period.

Data Storage

Storing data for the local program

The HIV surveillance program should develop protocols and procedures related to secure storage of program data in alignment with CDC's *Data Security and Confidentiality Guidelines* (available at https://www.cdc.gov/program-collaboration-service-integration/php/data-security/index.html) [2].

Physical storage at a minimum should consider the activities outlined in Guiding Principle 8, Guiding Principle 9, Standard 4.1, Standard 4.3, and Standard 4.4. These include "a dedicated, secure room would be the preferred location for confidential, secured data stored in locked, immobile cabinets"; "Policies should address retention of paper copies. If electronic copies exist, paper copies can be destroyed when no longer needed, in accordance with established health department policies"; and "Data security policies and procedures should specify who has access to secure areas."

Electronic storage at a minimum should consider the activities outlined in Standard 5.1, Standard 5.2, and Standard 5.4. These include "ensure that analysis data sets that can be accessed from outside the

secure area are stored with protective software and verify removal of all personal identifiers", "ensure that any electronic transfer of data is approved by the ORP and subject to access controls, and that identifiable data are encrypted before being transferred", and "use encryption software that meets federal Advanced Encryption Standard (AES) standards to encrypt data."

Additional Documentation

Data Management Process Modifications

Data management process modifications (e.g., data collection, additions, changes, deletions) should be documented, including the name of the staff member performing the action, the date, and the action. This can be helpful in monitoring how and why a case record has been changed over time. Most such changes are automatically documented in the surveillance information system and the various systems that support surveillance activities.

Additionally, it is especially important to create a record of all major programmatic changes related to statute changes, implementation of new data systems, case report form changes, case definition calculations, etc. Maintaining a record such as this can assist when describing points in the program's history when caution should be used in interpreting data trends, reporting volume, reporting delays, and other variations within the data.

Data Governance Documentation

The data governance documentation should provide an overview of the data management processes, procedures, and activities. At a minimum, the documentation should include:

- Data management staff organization chart
- Description of the roles and responsibilities of the data manager and each member of the data management team
- Timetable for each activity
- Contact information for relevant personnel outside the HIV surveillance program (e.g., contact person at the jurisdiction's vital statistics office)
- Description of the surveillance information system and how it interacts with other IT systems
- Description of how information flows into the surveillance information system and where the data are maintained

A **flow diagram** should be developed to help staff understand the various reporting sources and how data are received and entered in the HIV surveillance system. The document should describe

- Where and how data (both paper and electronic) are physically stored and archived
- Paper-based file-naming conventions
- Server locations and names

The document may also include data and data system architecture models.

Database Management Documentation

Database management documentation should include

- Description of the surveillance information system that is used to manage HIV surveillance data and the software manuals and documentation that accompany the software
- Relational database table diagram of the surveillance information system
- Plans for monitoring and improving database performance and capacity
- Descriptions of database security and access controls
- Description of the maintenance activities, such as backup (frequency) and restore

Data Security Documentation

Data security documentation [2] should include

- Description of the measures taken to ensure that HIV surveillance data are managed in a secure environment and that data are kept safe from unauthorized access and corruption
- Processes to describe how access to the database is monitored, including
  - procedures for requesting access to the surveillance information system
  - current log of persons who have access.

Data Quality Documentation

Data quality documentation should include descriptions of quality assurance activities, such as

- Routine data quality checks especially prior to data analysis and/or data transfer to CDC
- The routine follow-up of case records that contain errors
- Cases for which required information is missing
- eHARS matches that were conducted to improve data quality and completeness (see Technical Guidance file *Evaluation and Data Quality*)

Reference Documents

Reference documents should include

- Descriptions of standard naming conventions for files, folders, servers, databases, and documents
- Description of coding standards, such as
  - reference tables (e.g., facility and provider lists)
  - laboratory test codes (e.g., the Logical Observation Identifiers Names and Codes [LOINC]) that have been received by the HIV surveillance program.
  - Census or geographic information such as census tract, county FIPS code, ZIP codes, etc.

Other Documentation

Other documentation can include

- Steps for analytic processes
- Frequency of analytic data set creation
- Analytic data set retention (i.e., create eHARS SAS analytic data sets quarterly and save for 10 years)
- List of surveillance products and evaluation reports (and their timelines) that are generated by the HIV surveillance program routinely or ad hoc, such as
  - monthly summary
  - bulletins
  - annual surveillance report
  - fact sheets
  - slide sets
  - interim and annual progress reports
- Standardized method for analysis, such as which analytic data set should be used and variables to be used
- Availability of standardized SAS programs
- Any other type of analysis that involves disseminating data and evaluating data quality and system performance.

Records Management Documentation

Records management documentation should include

- Policies regarding the management of records (records can be paper [e.g., case report forms, laboratory reports] or electronic [e.g., archived data sets, end-of-year frozen data sets, lab reports])
- Description of how the information is stored, retained, archived, and destroyed (in keeping with state, local, and territorial laws and regulations)
- Schedule on which these procedures (preceding bullet) take place

Creating Data Management Reports

In addition, data management reports should be routinely created and shared with the appropriate parties:

- Reports of database and surveillance operations
  - Staff access to data, including read-only access and removal of staff who no longer need access
  - Changes/deletions
  - Workflow (e.g., cases by process completed, such as visual editing, follow-up investigations)

- Reports of cases by source of initial report
- Timeliness of reporting and the completeness of the data elements listed by each report source for feedback purposes
- Reports on quality control operations
  - Cases containing errors: eHARS records that have met the case definition but have a Person View status of E—Error, indicating that specific Person View edit-check rules have failed
  - Cases for which fields required by eHARS are missing: eHARS records that have met the case definition but have a Person View status of R—Required Fields Missing, indicating that one or more data elements that are required to report a case of HIV infection to CDC (last name Soundex, date of birth, sex, race, ethnicity, vital status, primary identifier assignment [i.e., STATENO] is missing, or date of death if vital status is "dead")
  - Cases in the surveillance information system that have not met the surveillance case definition for HIV infection
  - Monitoring unusual changes in aggregate case counts
  - Cases missing county FIPS of residence at diagnosis (information is used for metropolitan statistical area [MSA] assignment at the national level and used by CDC for reporting to other federal agencies)
  - Cases missing data elements that are important for state, territorial, local, and CDC analyses (e.g., type of facility at HIV diagnosis or stage 3 [AIDS] diagnosis)
  - Cases with data anomalies (e.g., unusually old age at diagnosis, stage 3 [AIDS] diagnosed years after death, unusual specimen collection date, unusual test result and/or result unit)
  - Potential intrajurisdictional duplicate case reports (see Technical Guidance file *Duplicate Review*)
  - Evaluation programs (CDC-developed SAS programs such as Evaluation of NHSS and Monitor RIDR Progress) (see Technical Guidance file *Evaluation and Data Quality*)

References

1. CDC. eHARS Technical Reference Guide. https://cdcpartners.sharepoint.com/sites/NCHHSTP/HSB/default.aspx. Updated July 2023. Accessed July 26, 2023. Please refer to the most current version of the document. [All topics are also available in eHARS Help.]
2. CDC. Data Security and Confidentiality Guidelines for HIV, Viral Hepatitis, Sexually Transmitted Disease, and Tuberculosis Programs: Standards to Facilitate Sharing and Use of Surveillance Data for Public Health Action. https://www.cdc.gov/program-collaboration-service-integration/php/data-security/index.html. Published 2011. Accessed May 2, 2023.
3. CDC. Introduction to eHARS, the HIV/AIDS Reporting System: User Guide/Student Manual. https://cdcpartners.sharepoint.com/sites/NCHHSTP/HSB/default.aspx. Updated July 2023. Accessed July 26, 2023. Please refer to the most current version of the document.

Appendix A

HIV Surveillance Primary Identifiers (e.g., STATENO or CITYNO) Assignment

In the enhanced HIV/AIDS Reporting System (eHARS), all cases meeting an HIV case definition are assigned a primary identifier for reporting to CDC. For state and territorial health departments, the primary identifier is referred to as the STATENO. For cities CDC provides technical assistance and support for conducting HIV surveillance, it is referred to as the CITYNO. The primary identifier is created and assigned by the site entering the case. Because names and other identifying information are not transmitted to CDC, the primary identifier is used nationally to distinguish a case of HIV infection and is used when CDC needs to communicate with the reporting site (i.e., provide information or ask questions about the case).

The identifier should not contain information that can be used to identify the person (e.g., name, residence), the reporting health department, or the testing site.

Each HIV surveillance program must work with its local health jurisdictions to ensure that primary identifiers are managed consistently. Multiple primary identifiers cannot be assigned to the same case in eHARS. However, many sites have external laboratory and HIV data management systems. In these systems, it is imperative that the case report of a person with diagnosed HIV infection be assigned one primary identifier and that deduplication be performed to ensure that one case report (regarding one person) is not assigned multiple primary identifiers (e.g., STATENO, CITYNO). Primary identifiers should be assigned only after thorough case ascertainment has determined that the case has not been previously reported and previously assigned a primary identifier.

If deduplication determines that a case report has been assigned more than one primary identifier (STATENO or CITYNO), the primary identifier that was entered first should be maintained, and the other, erroneously assigned primary identifier should be retired and never used again. Refer to the Technical Guidance file *Adult HIV Confidential Case Report Form* for additional guidance about assigning primary identifiers. The primary identifiers are considered sensitive, but not personally identifiable information, because they can be used to easily link to all direct identifiers and might be used to infer HIV status. See *Data Security and Confidentiality Guidelines* (available at https://www.cdc.gov/nchhstp/programintegration/Data-Security.htm).

Appendix B

Managing and Maintaining Facility and Provider Data in a Document-based Surveillance Information System

Collecting, managing, and maintaining complete and accurate data on the facility and the health care provider who diagnosed HIV infection or who has provided or is providing care to the individual with HIV are important HIV surveillance activities. Facility and health care provider information is essential not only for supporting Data-to-Care prevention and care activities and facilitating the collection of other HIV surveillance data, including supplemental HIV surveillance data [e.g., the Medical Monitoring Project (MMP)], but also for enabling facility-based analyses, identifying facility-based reporting differences, and improving the fit of statistical models.

As the reporting and importation of data in the surveillance information system becomes increasingly streamlined and automated, more ongoing attentiveness and diligence are required to manage the facility and health care provider data. HIV surveillance programs should have written Standard Operating Procedures (SOPs) for collecting and managing the data.

HIV surveillance collects the following data on facilities and health care providers:

- Facility:
  - Name
  - Address (street, city, county, state, zip code, country)
  - Email
  - Telephone number
  - Fax number
  - Email
  - Type of facility
  - Facility setting
  - Does facility receive funding from the Health Resources and Services Administration (HRSA)
  - Type of HRSA funding received
  - Unique identifier (auto-assigned by eHARS)
- Health care provider:
  - Name
  - Telephone number
  - Specialty
  - Unique identifier (auto-assigned by eHARS)

In accordance with the Centers for Disease Control and Prevention (CDC) Assurance of Confidentiality, facility street address, zip code, email, telephone number, fax number and health care provider name and telephone number are not reported to NHSS at CDC.

This document provides recommendations and suggests some best practices for managing and maintaining facility and health care provider data in a document-based surveillance information system.

Facility/Health Care Provider Creation in the Surveillance Information System

Before adding a new facility or health care provider in the surveillance information system, surveillance staff should first determine if the facility or health care provider is already known to the system. One may use the facility or health care provider's name, address, and/or telephone number to perform a search in the system. In eHARS, this can be done by using the ADMIN/Health Care Facility Maintenance or the Admin/Health Care Provider Maintenance utility.

If the facility or health care provider on the incoming surveillance report does not already exist in the surveillance information system, (i.e., no match exists when comparing the name, address, telephone number, and type) a new entry for the facility and/or health care provider should be created in the system. In eHARS, this can be done by using Add New, available under the Admin/Health Care Facility Maintenance and the Admin/Health Care Provider Maintenance utilities.

HIV surveillance programs should verify the completeness and accuracy of the facility and/or health care provider information being entered in the surveillance information system. This may be done by using a variety of methods and sources, including:

- Contact the facility or health care provider directly.
- Contact HIV prevention and care program staff who may be more familiar with the HIV care providers in the area.
- The National Provider Identifier (NPI) Registry, available at https://npiregistry.cms.hhs.gov/. The registry enables the user to search for a facility or a health care provider in the National Plan & Provider Enumeration System (NPPES). The system is maintained by the Centers for Medicare & Medicaid Services (CMS). There is no charge to use the registry.
- Other Web-based search services that may provide up-to-date data on a facility or health care provider's name, location(s), contact information, licenses, group/hospital affiliations, status and more. HIV surveillance programs must ensure that the Web-based search is conducted as securely as possible.

If the facility is affiliated with or is owned/operated by a larger institution (e.g., University of Washington Medicine owns a number of hospitals/medical centers and neighborhood clinics), HIV surveillance programs should have well-defined rules and instructions for when it is appropriate to create a separate entry for the facility in the surveillance information system. Factors to consider may include: Does the facility operate independently? Does the facility have its own name? Is the facility in a separate location? Are the types of care and services offered by the facility unique? If the facility's name is the same as the health care provider's name (e.g., Jane Doe's Adult HIV Care Center), a facility as well as a health care provider should be created in the surveillance information system. More specifically, in eHARS:

- Creating a facility allows the capturing of the facility's address, email, and fax number in the FACILITY_CODE table (since this information cannot be captured by the PROVIDER CODE table).
- Creating a health care provider allows the HIV surveillance program to capture the health care provider's name and specialty in the PROVIDER_CODE table (since this information should not or cannot be captured by the FACILITY_CODE table).
- To prevent inadvertent transmission of the name of the health care provider to CDC as well as to ensure the ability to link the entry in the FACILITY_CODE table with the entry in the

PROVIDER_CODE table, the facility name in the FACILITY_CODE table (FACILITY_CODE.name1) should equal the eHARS-assigned health care provider unique identifier in the PROVIDER_CODE table (PROVIDER_CODE.provider_uid).

Facility Naming Convention

DO NOT abbreviate a facility's name. The facility's name should be spelled out completely. For example, 'Grady Memorial Hospital' should not be abbreviated and stored as 'Grady' in the surveillance information system. Abbreviation hinders matching processes and increases the likelihood of false nonmatch. Historical data containing abbreviated facility names should be identified and cleaned up to enhance matching efficiency and accuracy.

For facilities with multiple locations, the naming convention may be 'facility main name' + '–' + 'facility sub name/city/town/area' (e.g., 'Planned Parenthood – Amarillo' and 'UC Davis Medical Group – El Grove').

For facilities that are located out of state, the naming convention may be 'OOS' + '–' + 'state abbreviation' (e.g., 'OOS – CA' and 'OOS – NH').

HIV surveillance programs should establish and follow a locally defined naming convention for facilities that should no longer be selected as the facility of diagnosis for persons newly diagnosed with HIV infection or stage 3 infection (AIDS) or as the current facility of care for persons receiving HIV clinical care (e.g., the facility is closed or shut down). For example, the name of such facilities may have an asterisk (*) prefix (e.g., FACILITY_CODE.name1 = '*Stewart-Webster Hospital').

Facility Type Assignment

There are presently 8 main facility type categories in eHARS:

1. Inpatient ('F01')
2. Outpatient ('F02')
3. Emergency room ('F03')
4. Screening/Diagnostic/Referral ('F04')
5. Laboratory ('F05')
6. Other specific facilities (e.g., correctional) ('F07')
7. Other ('F.OTH')
8. Unknown ('F.UNK')

A detailed description/definition for each of the main and subfacility type categories can be found in Chapter 4 (Lookup Codes) of the eHARS Technical Reference Guide. When assigning a facility type, HIV surveillance programs should refer to the descriptions/definitions and assign a type that best fits the type of care and services offered by the facility.

The following types of facilities are not presently accounted for in the facility type categories in the eHARS:

- Urgent care. HIV surveillance programs may assign such a facility a FACILITY_CODE.facility_type_cd = 'F03 – Emergency Room' in eHARS.

- Hospital clinic. HIV surveillance programs may assign such a clinic a FACILITY_CODE.facility_type_cd = 'F02.25 – Outpatient Facility/Other Clinic' in eHARS, as it is the best fit despite the description for F02.25, which suggests the type category may only be assigned to clinics with unspecified or unknown hospital affiliation.
- Telemedicine clinic. HIV surveillance programs may assign a facility offering only telemedicine services as FACILITY_CODE.facility_type_cd='F02.25– Outpatient Facility/Other Clinic' in eHARS through eHARS 4.13.1. For the facility short name (name2) in eHARS, consider including "Telemed" or "Telemedicine" so that this information can be used locally and at CDC to indicate whether the facility was a telemedicine company. Beginning in eHARS 4.14, use the facility type specific to a telemedicine clinic that will be created under the Outpatient Facility type. Please do not use this facility type to indicate when telemedicine is used as the method of service delivery by a facility that falls into other care and services categories (e.g., do not use for an infectious disease clinic that uses telemedicine for some patient appointments). There are not currently variables in eHARS to designate whether services were delivered during a specific encounter with a patient through telemedicine.

Although assigning a main type category (i.e., 'F01', 'F02', 'F03', 'F04', 'F05', 'F07') to a facility is generally sufficient, assigning it a more detailed subtype category when appropriate (e.g., 'F05.02 – Screening/Diagnostic/Referral: Drug Treatment Center') can provide beneficial and actionable information to HIV surveillance, prevention, and care programs, as well as for research and evaluation studies.

Facility Consolidation

Facility consolidation should be conducted, at a minimum, twice annually. If resources are available, consolidation should be done even more frequently (e.g., quarterly or monthly) to reduce the level of effort required during each consolidation and, more importantly, to improve facility search efficiency and maintain a high level of facility data quality in the surveillance information system.

To identify different facility entries in eHARS that may refer to the same facility (i.e., duplicate entries), HIV surveillance programs may run Auto Facility Consolidation or Manual Facility Consolidation, available under the Admin/Facility Consolidation utility. (The latter is recommended if a large number of facilities need to be merged.) When using Manual Facility Consolidation, surveillance staff may first run the Facility Consolidation Pre Report to get a listing of all facilities currently in the system. This list should be reviewed to determine if two or more facilities are the same facility. Matching criteria may be:

- Facility name AND
- Facility addresses AND
- Facility telephone number AND
- Facility type.

If two or more facilities match on name, address, telephone number and type, they should be merged. Merge the facilities in the eHARS test environment first and review the Facility Post-Consolidation Report. The report shows which facilities have been consolidated, the number of cases and documents that were affected by the consolidation and a line list of affected cases (ehars_uid and STATENO) and documents (document_uid and document_type_cd). If the process is successful in the test environment, staff may proceed to repeat the same facility consolidation tasks in the eHARS production environment.

In keeping with the practice of document-based surveillance, HIV surveillance programs should have well-defined rules and instructions for when two or more facility entries in the surveillance information system that refer to the same facility should remain as separate entries and not be consolidated. Examples may include:

- The facility was relocated, i.e., address changed. Keeping two separate entries for the facility in the surveillance information system allows HIV surveillance programs to retain the historical address data and examine trends in HIV service coverage over time.
- The facility type changed (e.g., from 'F04.05 – Screening/Diagnostic/Referral/HIV Counseling and Testing Site' to 'F02.51 – Outpatient Facility/Community Health Center'. Retaining the historical facility type data is important for statistical data analyses.

Surveillance staff may run the Facility Consolidation Pre-Report available under the Admin/Facility Consolidation utility in eHARS to identify facilities that have never been used (i.e., the number of associated documents for the facility is 0). There are two ways to "delete" a facility from eHARS if it has never been used:

- Contact the Division of HIV Prevention (DHP) Helpdesk at 877.659.7725 or nchhstpinformatics@cdc.gov or dhpsupport@cdc.gov to request a correction script.
- Create a dummy facility in eHARS. Each time a facility is determined to have never been used, consolidate it with the dummy facility by using the ADMIN/Facility Consolidation utility.