STI and HIV Data Linkage Manual:

Improving Ryan White HIV/AIDS Program Outcomes using a Tiered Technical System Approach to HIV-STI Data Linkages

GEORGETOWN UNIVERSITY

**Center for Global Health
Practice and Impact**

Table of Contents

Acknowledgement

Abbreviations

ADPH	Alabama Department of Health
BPD	Business Process Development
CDC	Centers for Disease Control
CGHPI	Georgetown University's Center for Global Health Practice and Impact
CoP	Communities of Practices
DC SID	Washington DC's Strategic Information Division
DOB	Date of birth
DOH	Department of Health
DSA	Data Sharing Agreement
DWH	Data warehouse
D2C	Data to care
eHARS	Enhanced HIV/AIDS Reporting System
ETL	Extract, transfer, and load
FDOH	Florida Department of Health
HIV	Human Immunodeficiency Virus
HRSA	Health Resources Services Administration
IOP	Internal Operating Procedure
KII	Key Informant Interview
LDH	Louisiana Department of Health
OOC	Out of Care
OOJ	Out of Jurisdiction
QA	Quality assurance
RWHAP	Ryan White HIV/AIDS Program
STFP	Secure File Transfer Protocol
SHHP	STD/HIV/Hepatitis Program
SOP	Standard operating procedure
SSH	Secure Shell Protocol
SSN	Social Security number
STI	Sexually Transmitted Infection
TA	Technical assistance
TWG	Technical Working Group
VL	Viral load

Introduction

The Health Resources and Services Administration (HRSA) Ryan White HIV/AIDS Program (RWHAP) Part F Special Projects of National Significance (SPNS) HIV-Sexually Transmitted Infection (STI) Data Linkages project's purpose is to provide technical assistance (TA) to jurisdictions in communities and states that improve HIV and STI data linkage and coordination to improve data utility for people with HIV among RWHAP clients.

The synchronization of STI and HIV surveillance data will be used to improve the capacity of health departments and RWHAP providers to prioritize resources for linking and re-engaging people with HIV into care. Improved data linkage and synchronization will lead to more data accuracy and will ultimately inform a timely review of client-level care status. Data accuracy and synchronization have been documented to lead to better care outcome activities, such as increased linkage to care, increased engagement and re-engagement in care, and ultimately, increased viral suppression among people with HIV. Enhanced synchronization of disparate data systems assists in confirming residence and care patterns among people diagnosed with HIV, including people lost to follow-up and/or in care in other jurisdictions. These efforts are critical for re-engagement and other client-level intervention activities. Enhanced data synchronization also increases the ability of public health departments and HIV providers to assess risk behaviors, repeat STI infections, and other critical markers of relevance in addressing co-infections of STIs in people with HIV.

Four jurisdictions participated in the HRSA 19-039 SPNS project "Improving Ryan White HIV/AIDS Program Outcomes Using a Tiered Technical Assistance Approach to HIV/STI Data Linkages:" Florida, Alabama, Washington, D.C., and Louisiana. Throughout this project, the Florida Department of Health (FDOH) enhanced and automated a centralized HIV and STI data warehouse, coordinated external systems and processes, and now shares data among six independent country-level programs. The Alabama Department of Health (ADPH) also developed a centralized HIV and STI data warehouse, as well as internal dashboards to inform prevention and outreach while improving collaboration between the ADPH, RWHAP, and county-level providers. DC Health improved business process documentation, developed routine data sharing between surveillance and patient care, and established coordinated data to care activities between divisions. The Louisiana Department of Health (LDH) integrated HIV data into current STI programs, formalized all departmental data processes, and enhanced care data accessibility within the program's care outreach system.

This manual has three major sections: Design, Implementation, and Evaluation. Each section includes tools and guides to facilitate STI and HIV data linkage and utilizing data to engage people with HIV in care. The final section of the manual includes templates that can be modified per specific project needs and objectives.

Design Guides and Tools

<u>Tool 1: Needs Assessment</u>

Needs Assessment Tool Contents

This tool contains the following five areas for the preliminary needs assessment:

1. **Jurisdiction Readiness**
   a. Level of Integration between Core Teams
   b. Staffing of Core Teams
   c. Current Data Sharing Policies

2. **Data Integration Activities: HIV and STI**
   a. Goals and Concerns regarding Linkage of HIV and STI Data
   b. Existing Technology used for HIV/STI Surveillance Data
   c. Existing Processes for HIV/STI Surveillance Data

3. **Ryan White HIV/AIDS Program**
   a. Ryan White HIV/AIDS Program Data

4. **Data-to-Care Activities**

5. **Pathway to Data Integration**

Background
The Needs Assessment should be structured as a questionnaire to be sent out to project stakeholders to assess the current state of HIV/STI linkages and RWHAP/Data to Care processes, and to understand your project's goals and needs moving forward. The information provided in questionnaire responses will be used to inform the Data Integration Implementation Plan, which will serve as a project roadmap.

While this tool provides a foundation for your Needs Assessment, it is important to tailor questions to the specific context of your organization. The results of this assessment will help to identify existing gaps in your processes and inform your project's priority areas.

🏛 Jurisdiction Readiness

I.   Level of Integration between Core Teams

First, it is important to get an understanding of the organizational structure of the "core teams" within your organization to understand the level of structural integration within your jurisdiction's department of health. These teams may include but are not limited to: HIV Surveillance and Epidemiology teams, STI Surveillance and Epidemiology teams, RWHAP team(s), and Data to Care team.

Explain the organizational structure of the core teams listed above.	*For example, are the HIV/STI surveillance programs within the same overarching program as RWHAP, etc.*
Explain the level of collaboration and coordination that currently exists between each team.	*Use the table below to further quantify your thoughts on the effectiveness of communication/collaboration between each team.*

	No collaboration	→	→	→	Extremely effective/regular collaboration	Fully integrated team
	1	2	3	4	5	N/A
HIV and STI Surveillance						
RWHAP & HIV						
RWHAP & STI Section						

Data to Care/Client Re-engagement and HIV section					
Data to Care/Client Re-engagement and STI section					

II.   Staffing of Core Teams

Please complete the following chart regarding your staffing for each core team as listed below. If one of your teams is fully integrated with another core team listed below, please indicate in the comment box, and list N/A for the second of these teams.

	HIV Surveillance Team	STI Surveillance Team	RWHAP Team	Data to Care Team
STI Data Manager				
Data Entry Staff				
Senior Level Epidemiologist				
Bioinformatician/ IT Specialist				
Doctoral/Masters Level Services Planner				
Clinical Quality Improvement Specialists				
Disease Intervention				

Specialists				
Other (*explain*)				
Explain any N/A and 'other' from the chart above.				

III. Current Data Sharing Policies

Is a data sharing agreement or data use agreement needed to share between your jurisdiction's STI and HIV data systems?	*Yes or No? Explain further if needed.*
Is a data sharing agreement or data use agreement needed to share between your jurisdiction's HIV and Ryan White HIV/AIDS Program data systems?	*Yes or No? Explain further if needed.*
Is a data sharing agreement or data use agreement needed to share between your jurisdiction's STI and Ryan White HIV/AIDS Program data systems?	*Yes or No? Explain further if needed.*
Are there any data-sharing agreements currently in place between your jurisdiction's core teams?	*If yes, please explain current data-sharing agreements between your jurisdictions' core teams (HIV & STI Surveillance, Data-to-Care) including the scope of the agreement, involved parties, time frame of agreement, etc.*

Data Integration Activities: HIV and STI

I. Goals and Concerns regarding Linkage of HIV and STI Data

What are your jurisdiction's goals/rationales for linking HIV and STI data? *Select all that apply.*
A. Improved partner services activities
B. Creation of comprehensive epidemiological profile and Integrated Plans
C. Enhanced data-to-care activities
D. Improved HIV/STI surveillance data quality
E. Other (explain in text box below)

Explain any N/A or 'other'	*Use this text box to elaborate on goals above/explain additional goals as needed.*

II. Existing Technology Used for HIV/STI Surveillance Data

Is there a data governance model in place for data sharing and integration?	*What are the current policies in place to manage HIV and STI data? If yes, list the governance board.*
Where are your **HIV** surveillance data housed and/or what software system do you use (i.e., eHARS, MAVEN, etc.)?	*Provide a brief description.*

*Please rate your experience using this software (for **HIV** data) in the table below:*

	Poor	→	→	→	Excellent
	1	**2**	**3**	**4**	**5**
Ease of Use					
Reliability					
Utility					
Cost					
Time-Intensiveness					
Where are your **STI** data housed and/or what software system do you use?	*Provide a brief description.*				

*Please rate your experience using this software (for **STI** data) in the table below:*

	Poor	→	→	→	Excellent
	1	**2**	**3**	**4**	**5**
Ease of Use					
Reliability					
Utility					
Cost					
Time-Intensiveness					

What methods are you using to link HIV and STI data (if any)?	*Provide a brief description.*

*Please rate your experience using these methods (for **HIV/STI data linkage**) in the table below:*

	Poor	→	→	→	*Excellent*
	1	**2**	**3**	**4**	**5**
Ease of Use					
Reliability					
Utility					
Cost					
Time-Intensiveness					

III.   Existing Processes for HIV/STI Surveillance Data

How are HIV and STI data initially entered and verified?	
Explain measures in place to check the validity and reliability of HIV and STI data.	
Are there processes in place to link HIV and STI data?	*If so, explain these processes.*

Does your jurisdiction link HIV and STI data regularly? If so, how often?	

Does your jurisdiction have current methods in place to link HIV and STI data?	*If so, please explain this process and the issues that you face.*
What are some of the challenges you face in linking HIV and STI data?	
Please use this space to discuss any other questions, concerns, or comments regarding HIV/STI data linkage.	

Ryan White HIV/AIDS Program

Briefly explains how better integration of HIV and STI data will assist with improving the HIV care continuum within the RWHAP.	
How are RWHAP Parts A or B data collected?	
How are RWHAP Parts A or B data entered?	
Where are your RWHAP Parts A or B data housed and/or what software system do you use?	

*Please rate your experience using this software (for **Ryan White HIV/AIDS Program** data) in the table below:*

	Poor	→	→	→	Excellent
	1	2	3	4	5
Ease of Use					
Reliability					
Utility					
Cost					
Time-Intensiveness					
What					

verification checks are in place for your RWHAP data post-data entry?	
Is there data sharing between the HIV/STI departments and the RWHAP databases?	*Please elaborate*

*Please rank the **quality** of your jurisdiction's RWHAP data:*

	Poor	→	→	→	*Excellent*
	1	**2**	**3**	**4**	**5**
Completeness					
Readability					
Utility for data-to-care					
Use this text box to elaborate on concerns with RWHAP data quality.					

Data-to-Care Activities

	Explain with a focus on data utilization and gaps.
Explain your jurisdiction's data-to-care activities with regard to the RWHAP clients served and the HIV care continuum.	
Explain the various pathways for client care re-engagement in your jurisdiction. Do you conduct care matches with providers? With Part A programs? Do you have a data-to-care program at the Department of Health (DOH)?	
Elaborate on any challenges your data-to-care team faces with data quality, logistics, capacity, or client re-engagement.	

Pathways to Data Integration

Please quantify the following potential concerns that your jurisdiction may have with the process of integrating HIV/STI data:

	Of less concern	→	→	→	*Of most concern*
	1	**2**	**3**	**4**	**5**
Quality of HIV surveillance data					
Effectiveness of HIV surveillance data system					
Quality of STI data					
Effectiveness of STI data system					
Capacity to regularly link HIV and STI person-level data					
Capacity to store linked HIV and STI data within a single system					
Availability of integrated HIV and STI data to DIS and outreach staff					

Quality of linked HIV and STI data				
Capacity to perform data quality matches between HIV and STI data and other DOH data systems (ADAP, Pharmacy, etc.)				
Developing data sharing agreements between DOH and RWHAP providers				
Capacity to perform data sharing between DOH and RWHAP providers				
Technological capacity for data linkage between DOH and RWHAP provider systems				
Efficacy of data sharing with RWHAP providers in returning and retaining clients in care				
Staffing capacity to maintain continued data integration practices				
Availability of documentation related to all DOH data systems				
Availability of documentation				

related to all data exchange and data integration procedures				
Availability of documentation related to data quality and validation procedures				
	Use this text box to elaborate on the above as needed.			
Explain any N/A or 'other'				

Conclusion

Please use this space to enter any concerns or additional needs that you feel were not covered in this questionnaire. Note that this project should continuously assess the needs throughout implementation. This questionnaire is meant only to provide an in-depth, first-look assessment of your jurisdiction's needs.

Tool 2: Post-Assessment Guide

Use this guide following the conclusion of your preliminary needs assessment to translate information collected into the foundation of your project implementation plan. Completing the exercises outlined in this document will help you to reflect on the outcomes of your needs assessment and begin to define the primary priority areas of your specific implementation. You should plan to use this document.

Identify Priority Areas for Action

Use the findings of your preliminary needs assessment to determine your organization's strengths and weaknesses as they correspond to each section of the survey. Using the table provided below, identify the strengths and weaknesses of each survey area.

Survey Section	Strengths	Weaknesses
Jurisdiction Readiness		
Data Integration Activities: HIV and STI		
Ryan White HIV/AIDS Program		
Data-to-Care Activities		
Pathway to Data Integration		

Next, take the identified weaknesses and categorize them into four or more core thematic areas. What are the most pervasive themes found in the analysis of weaknesses? These thematic areas will become your priority areas. Use the table below to organize these themes.

Example Priority Area	*For example, you have identified that your organization operates in siloed departments. Standardized communication between two or more entities (i.e. HIV and STI) is lacking. Additionally, outdated Internal Operating Procedures (IOPs) lack clarity and contribute to poor communication structures. The overarching theme identified here involves Communication. This thematic area can become Priority Area 1.*
Priority Area 1	
Priority Area 2	
Priority Area 3	
Priority Area 4	
Priority Area 5	

Root Cause Analysis of Priority Areas

For each priority area, perform a root cause analysis on why these areas are weaknesses. This is a crucial step for better understanding current processes and existing gaps in your work.

Example of a Root Cause Analysis:

Problem: The communication between internal departments is poor.

→ Why? Tensions exist between departments.

    → Why? Duplication of efforts exists due to a lack of information sharing.

        →Why? There is lacking clarity as to what data and information can be shared between departments.

            → Why? No clear documentation exists on proper communication flows.

Solution: Establish/update Internal Operating Procedure (IOP) for communication.

Additionally, it is helpful to fill in essential information for each priority area to inform your implementation development. The following sections in the table below provide some *examples* to get you started.

Resources and Staffing	Summarize current resource and/or staffing restraints; Outline gaps in staffing and/or resources; Propose next steps.
Data-to-Care Constraints	Summarize current data-to-care capacity issues; Describe current data-to-care activities; Outline roadblocks for data-to-care capacity; Propose next steps.

Data Sharing Constraints	Summarize the existing logistical/legal constraints on data sharing within your jurisdiction; Outline the organizational structure of your jurisdiction departments with regards to data sharing; Outline existing roadblocks to data sharing; List any existing Data Sharing Agreements (DSA) within departments; If DSAs are needed- between which departments and to what specifications; Propose next steps.
Ryan White HIV/AIDS Program Data Quality	Summarize your current capacity constraints for RWHAP data quality; Describe current methods for RWHAP data collection and entry; Detail current software used for RWHAP data; Describe current verification checks for RWHAP data; Propose next steps.
HIV or STI Data Quality and/or Software Issues	Summarize HIV/STI data quality/software issues; Outline techniques currently used for data quality assurance; List the software used to manage HIV/STI data; Outline key components of HIV data quality that your jurisdiction currently struggles with; Propose next steps.
HIV/STI Linkage Methods	Summarize HIV/STI data linkage issues; Describe your current methods for HIV/STI data linkage; Describe the current frequency of data linkage activities; Describe the current verification checks for HIV/STI data linkage; Outline the key issues of HIV/STI data linkage capacity; Propose next steps.

Next Steps

Now that you have completed the preliminary needs assessment and defined your priority areas, you can begin developing specialized working groups to address the objectives you have set forth in this document. The next document, the *Technical Working Group (TWG) Development Guide*, will provide guidance on how to move from planning to implementation. This is where you turn your objectives into action.

Tool 3: Technical Working Group Development Guide

This document will be broken down into three distinct sections:

1. Objectives into Action
2. Applying and Implementing Technical Working Groups
3. Maintaining Technical Working Groups

Objectives into Action

Context: Once the priority areas have been identified, **Technical Working Groups** (TWGs) can be utilized to focus specialized resources and activities toward accomplishing specific objectives. TWGs make up a smaller subset of the overall project where participants contribute their expertise on a specific topic or problem. The goal of TWGs is to foster close collaboration amongst key stakeholders and members dedicated to devising actionable solutions. This collaboration is a proven method for achieving desired outcomes through efficient use of time and available resources.

Applying and Implementing TWGs

When applying TWGs to your project, it is important to tailor all solutions to the specific needs established in your preliminary Needs Assessment and your Post-Assessment analysis. The following six steps provide a general framework for your TWG development.

Six-Step Framework for TWG Development:

1. Describe the problem	Provide sufficient background information to detail the selected problem. Be sure to include the current challenges, goals for improvement, and desired outcomes.
2. Identify the experts	Engage in-house experts to maximize effective use of resources and develop localized solutions.
3. Documentation	Ensure documentation throughout the TWG lifespan to track progress and changes made in the process.
4. Define the scope of work	Once the participants have been selected, work together to determine the group's scope of work, including group expectations and deliverables. This step is essential for setting baseline targets and ensuring all members are on the same page.
5. Delegate responsibilities	Assigning roles and responsibilities ensures accountability and clear expectations of involved parties.
6. Establish a schedule	Collectively determine a timeline of deliverables and the frequency of group meetings needed to achieve desired outcomes.

Maintaining TWGs

The final essential piece of this guide involves active and frequent adjustments throughout the TWG's progress. Allowing the flexibility of course correction makes space for your TWG to practice continuous improvement. This guide should serve as a live document, updated regularly to track progress and document any necessary changes. Use the table below to document progress, updates, and changes to each of your TWGs.

Date	Progress	Changes	Comments

Implementation Guides

Implementation Guide 1: Staffing

Staffing is one of the most integral parts to successfully implement the Data Linkages project. Having a solid plan helps the organization to achieve personnel needs as well as resources. It also allows the opportunity to identify gaps and determine what outside resources and additional staff are needed. Because of this, it is important to have the right people sitting at the table and involved in the process. Ideally, this process would be a collaborative effort for all programs and/or departments involved. Therefore, there should be representatives from each program and/or department. When working on implementation, keep in mind that staffing plans will be different for each organization based on several aspects such as resources, funding, and size along with other factors. Below are some suggestions as to the different departments and personnel who should be involved in this data linkage process.

Leadership
The leadership team should be involved as they are the ones who will give the final stamps of approval necessary to get the activities started. They may also be the ones who help to bring all parties to the table and facilitate conversations to convey the importance of the work.

HIV Surveillance Staff
HIV Surveillance staff should be at the table because the data from this department will be the basis of the integration process. From this department staff such as:
- The Department Lead – To provide necessary approvals and guidance to ensure a smooth process.
- Epidemiologist – To work with the other departments to prepare for the data linkage.

STI Surveillance Staff
The STI surveillance staff should be involved because the nature of this work is to link STI and HIV data. Similar to HIV surveillance, this team should be inclusive of:
- The Department Lead – To provide necessary approvals and guidance to ensure a smooth process.
- Epidemiologist – To work with the other departments to prepare for the data linkage.
- Disease Intervention Specialist (DIS) Supervisor and/or DIS – To provide feedback on how helpful the data linkage has been since it has been completed.

Ryan White HIV/AIDS Program Staff
The goal of the data linkages is to help improve HIV and STI outcomes among clients, therefore when and where RWHAP data can be used, it should be to add to the completion of the data view. Staff from the RWHAP team should be:
- The Department Lead – To provide guidance and oversight in the linkages process.
- Data Analyst – To work alongside epidemiologist to pull the necessary variables needed for linkages.
- Ryan White HIV/AIDS Program Case Manager – To provide feedback from the user endpoint about what data is necessary to provide a complete client view.

IT Staff

IT staff are integral for this process as they can handle all logistics that fall outside of the scope of the surveillance staff. They can help with data management and storage, depending on how many IT staff are available within each individual department or within the whole DOH.

Implementation Guide 2: Best Practices for Business Process Development

Introduction

Business Process Development (BPD) is a way of looking at and improving the inner workings of an organization to make them more efficient and cost effective. When working in HIV and STI surveillance, you are often working with health departments that have limited time, resources, and/or staff. Streamlining business processes helps utilize limited resources (time and effort) and ensure all members within a single department and across departments have access to baseline information about how things work. In HIV surveillance, these activities help to maintain a comprehensive understanding of data sources, uses, reporting requirements, and data use practices. It strengthens surveillance units by minimizing duplicative efforts and inefficient processes, and creating a strong business process understanding to help maintain data-driven intervention designs with continuous quality improvement.

Business Process Overview

There is a breadth of literature related to business process development, but the following key steps are adapted from the Association for Intelligent Information Management page on business process management:
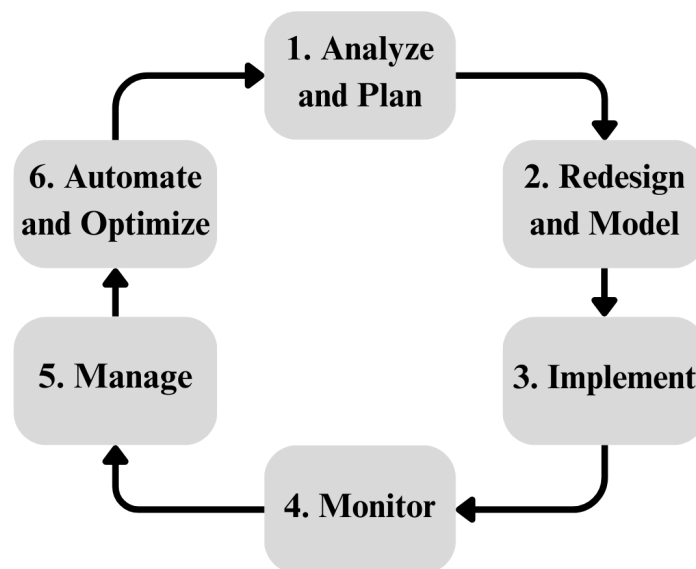
Workflow for Business Process Management

These steps are briefly described below:

1. **Analyze & Plan**
   In this initial step, the goal is to understand the current business process, identify the 'ideal process,' and outline specific outcomes and aims for the overall process. It is important to gather current documentation and assess scheduling, time requirements, staff capacity, gaps in documentation, and gaps in process efficiency.

2. **Redesign and Model**
   Revisit the defined needs of the 'ideal process' and compare it to the current process. Quantify time and effort expenditure under current processes, design feasibility testing for all re-designed components, and develop monitoring and evaluation plans to assess implementations. Discuss and outline components of the redesign and model the proposed process changes using a diagram, narrative, or workflow chart.

3. **Implement**
   Carry out feasibility testing prior to large-scale process change implementations, and then implement the proposed process changes according to a predefined work plan and timeline.

4. **Monitor**
   Carry out the monitoring plan for all re-designed aspects of the business process and assess the need for further modifications by deciding which changes are cost-saving, do not affect the cost, and which are cost-depleting.

5. **Manage**
   Incorporate beneficial components of re-design into existing business processes by documenting all changes into a coherent and comprehensive updated business process. Document all monitoring and validation testing and train personnel as needed.

6. **Automate and Optimize**
   After the business process is optimized and accepted by all stakeholders, design automation techniques, feasibility testing, validation testing, and an implementation plan for applicable processes. Once processes are automated, evaluate the effectiveness and accuracy of automated processes, utilizing feedback as needed to optimize.

This process is cyclical and there should remain a continuous effort to optimize and improve processes over time.

Considerations for Business Process Development (BPD)

When determining ways to improve business processes, several things need to be considered. The first is available resources, time, and team buy-in. Understaffed or overburdened health departments could oftentimes benefit from BPD, but have little time or energy to dedicate to certain activities. The initial needs assessment for jurisdictions should evaluate the business process needs alongside any existing limitations. The goal of BPD is to create materials and processes that jurisdictions will actually use and benefit from — maintaining utility, convenience, and relevance to their work. Second, is staff turnover. As team members leave, process knowledge leaves with them. This leaves the remaining staff to train new staff members, oftentimes without support from written standard operating procedures (SOP) or other documentation. Oftentimes immediate needs overshadow the long-term need to document

procedures, and staff members do ad-hoc changes to business processes rather than dedicating time to mapping out these efforts ahead of time. As a TA provider, it can be helpful to assist departments to set apart dedicated time to coordinate these efforts across departments (which helps reduce duplicative efforts) and help overburdened staff with optimized BPD projects based on their most acute needs. Lastly, is to optimize BPD. Optimizing BDP can be performed by creating SOPs to document essential functions of a department or organization. SOPs can take the form of traditional written documents or could include narrated screen recordings showing how to navigate data systems or perform tasks needed for data linkage. Depending on the team's ability, creating a comprehensive written document might be too much of a burden on top of day-to-day responsibilities. Recorded procedures, meetings, and webinars can help relieve the burden of producing a long-form written document by a member of the staff while ensuring information is readily available through staff turnover.

Jurisdiction Business Project Development Outcomes

Georgetown University's Center for Global Health Practice and Impact (CGHPI) has taken a tiered and tailored TA approach to implementation design, due to the drastically different situations in each state. Each jurisdiction had a tailored data integration implementation plan as well as Technical Working Groups (TWGs) that met regularly and organized around the three focal areas. Additionally, to promote knowledge exchange and diffusion of knowledge and best practices, CGHPI created Communities of Practices (CoP) to engage participating and non-participating jurisdictions in critical knowledge exchange, best practices benchmarking, and collaborative learning. The major BPD outcomes in each jurisdiction are outlined below.

Alabama

The Alabama Department of Health (ADPH) has intense staffing limitations. CGHPI helped work with the staff to diagnose these limitations, reduce duplicative efforts, and help automate processes, when possible, to free up staff resources. ADPH has developed a centralized warehouse to integrate HIV and STI data and developed an internal dashboard to inform prevention and outreach that updates automatically to reduce the burden on non-coding trained staff. Through this, there has been improved collaboration among ADPH, the RWHAP network, and county-level providers.

Florida

At the beginning of the project, the Florida Department of Health (FDOH) had little communication between the STI and HIV staff, resulting in undefined responsibilities, duplicative efforts, and neglected tasks. All parties had to be brought together to better understand each other's roles and responsibilities so as to create effective processes for the design and use of an integrated data warehouse. FDOH enhanced and automated a centralized HIV-STI data warehouse, coordinated external systems and processes, coordinated strategic data sharing with six independent country-level programs, and completed a survey for HIV/STI groups, which is providing helpful feedback.

Louisiana

Louisiana Department of Health (LDH) processes had previously been done on an as-needs ad-hoc basis, and minimal documentation existed. Staff felt overwhelmed by data systems, transfers, and reports circulating with very little direction. LDH needed a better understanding of the processes and responsibilities within the STD/HIV/Hepatitis Program (SHHP) to develop clear business processes. LDH integrated HIV data into current STD programs, formalized all departmental data processes, and enhanced care data accessibility within the program's care outreach system.

Washington, D.C.

Washington, D.C.'s Strategic Information Division (SID) was overwhelmed by data systems, transfers, and reports circulating with very little direction. By establishing well-documented business processes and creating sustainable routes of communication, DC SID developed routine data sharing between surveillance and care, established coordination and synergy between departments with regard to routine data-to-care activities, and updated their data integration implementation plan.

Conclusion

Business process development can help streamline processes to make limited staff efforts more meaningful, standardize information within and across departments, and maintain knowledge throughout staff turnover. In health departments dealing with multiple conflicting priorities, creating seamless and automated processes (when possible) help make the most of staff members' time and energy while improving overall efficiency.

Implementation Guide 3: Best Practices for Data Integration

Introduction

Data Integration is combining related data from disparate sources to gain insights that may not have been evident without integrating data sources to ultimately translate data into meaningful actions. In HIV Surveillance, Data Integration leads to:

- **Increased collaboration** through the reduction of silos between different offices within the same Department of Health.

- **Better data quality** through the reduction of missing data, data discrepancies, missing insights, outdated data, and so on.

- **Improved outcomes** through data literacy, improved capacity to make data-driven decisions, and overall improved data to care (D2C) outcomes.

Baseline Investment in Collaboration and Scope-Setting
Bring all system purpose/vision documentation developed thus far before stakeholders (as defined by the jurisdictions) to:
- Receive feedback on the defined outcomes from stakeholders' perspective.
- Define purpose, roles, and responsibilities of stakeholders.
- Establish engagement and participation of stakeholders.
- Establish lines of communications between all stakeholders and related groups.
- Develop a communications plan for the data system, including a feedback loop.

Define Data Sharing Needs/Purpose
- Define data elements to share between departments.
- Define timelines.
- Develop work plan for design and implementation.
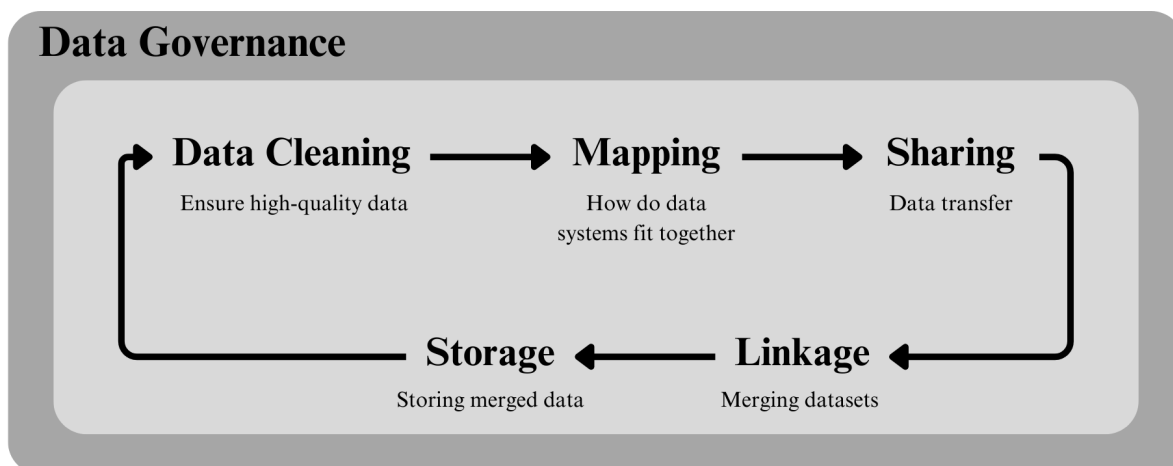
Data Integration Overview

Figure 1: Overview of Data Governance subsections.

Considerations for Data Sharing: Detailed Guidelines

Data Cleaning

Before integrating disparate data sources, it is important to assess the quality of each of your data sources. Sharing 'bad' data has the potential to exponentially increase the impact of those data, undermining the intended benefits of integrating data for improved data quality. Therefore, there are certain **quality assurance checks** that you should perform within each source system prior to moving forward with data exchange to exchange high-quality, reliable data, where limitations are identified and controlled beforehand.

Identifying data quality issues. There are a myriad of data quality issues that might arise in your data set, and some that you may not be able to control. It can be helpful to think of data quality issues in three categories:

1. Issues that you can identify, and control for.

2. Issues that you can identify, but cannot control for.

3. Issues that you cannot identify.

Each of these categories is important to have in mind, but there are limited steps you can take for the second and third categories. However, it is important to consider the uncontrollable data errors to ensure that all parties that have a stake in the data integration understand the potential that data are erroneous, so they can weigh the risks of making decisions based on the data with an understanding of the reliability of data. The subsequent sections will review descriptions, examples, and data quality assurance steps for each of these three categories.

1. For issues that you can identify, and control for. Sometimes there are data fields that are missing altogether or are filled but logically are not plausible, and these can be easily identified through a quick query of your dataset. Some examples are included in Table 1 below, along with remediation recommendations.

Table 1. Identifying and remediating common data quality issues

Data Error	Example	Identification Steps	Remediation Steps
Empty Data Fields	Date of birth (DOB) is missing for a client	Develop a list of 'mandatory data fields' or those that you would expect every entry in your dataset to have a value for. Then query these columns to calculate the number of rows with empty cells for these variables.	Ongoing quality assurance (QA) of missing data fields utilizing batch matching* to backfill/update person-level identifiers.
Conflicting dates	● Date of HIV Diagnosis is *before* DOB ● Labs dated after death dates	1. Develop an inventory of dates that you have in your data set. 2. Compile a list of rules that the dates should logically follow. 3. Query your data to identify instances where the dates for a single row are not following the rules.	After identifying erroneous date entries, you will need to manually review to see if there are any obvious errors (i.e., if the labs are coming in after a client has been marked dead, verify the source of the death information). Cross-check other data sources, and do manual look-ups to attempt to correct where possible. Where not possible, consider omitting erroneous data and sharing empty data rather than 'bad' data.
Conflicting geographic information / Recency of address information	Address of residence does not exist within the state of residence, not a real address, etc.	These will be harder to identify, but based on your level of validation, try to ensure that address information is logical - and that the data for each address field come from the same source. For example, if the state is listed as DC, but the zip code is a CA zip code, something is wrong.	Ongoing QA of address data utilizing batch matching* to backfill/update patient locating information.

*See the section on routine data quality assurance activities below for more information on batch matches.

2. For issues that you can identify but cannot control for. With every data set, there are limitations, and the best we can do sometimes is to acknowledge that those limitations exist, and ensure that when we share our data set, we communicate the known limitations clearly. Some examples of data quality issues that fall under this category are included in Table 2 below.

Table 2. Identifying and communicating the impact of common data quality issues

Data Issue	Example(s)	Potential Impact / Disclaimer
Missing Labs	We know that we do not have all of the HIV viral load (VL) labs due to previous assessments, due to having some labs that are still using paper-based reporting, or VLs that are taken as part of clinical studies and not transmitted to the health department.	Missing viral load labs may cause us to under-estimate care provision for HIV clients, or may lead to overstating the level of non-suppressed people with HIV in the region.
Contact Information Recency	If you do annual batch matches with a system like LexisNexis, ensure that you are transparent that dates may be up to a year old, and any further updates are limited to what is reported on incoming labs.	Oftentimes programs rely on contact information data for case management services to attempt to contact clients who have seemed to fall out of care. If feasible, include a 'date_of' data field attached to each contact field/address to provide recency context to the data end user.

3. For issues that you cannot identify. There will always be errors in the data that we cannot identify. For example, someone may fill out an incorrect Social Security number on a patient intake form that gets used for a case report form. Someone may have changed their name or gender identity, and that information may not have reached the data system yet. There are routine data quality assurance checks that you can do to keep your data as current and accurate as possible, but there will always be gaps. When considering the potential of unknown errors, the most important thing is to acknowledge that your data may be out of date and to regularly discuss efforts to both improve data quality and responsibly use data that are available despite unavoidable limitations.

Recommendations for Routine Data Quality Assurance Activities.

1. Don't wait for data exchange — do periodic checks on the issues discussed in this section! If you wait until you have a data exchange request to even probe your data for quality issues, you may find an unanticipated volume of data issues to review. The first attempt at system-wide data quality review may be overwhelming, but with recurring efforts, the workload will decrease each time.

2. Routine batch matches with national- and state-wide data systems to routinely quality assure associated data fields. See Table 3 for more information on common batch matches and associated data fields that can be updated with each.

3. Regularly assess process-related data quality errors. Ensure there is a pathway for staff to provide feedback on common errors that they find in the data and ensure that highly-reported errors are followed up on to identify the cause of the recurrent error. For example, if someone notices that the lab data suddenly has a common error in the VL result coding, try and identify where in the HL-7 process the error is occurring. Is the data coming through the HL-7 feed with the error, or is the error occurring as part of the ETL process at the health department? If it is arriving with the error, are the errors centralized within a certain lab company?

Table 3. Batch Match Systems and Associated Data Fields

System (linked)	Potential QA Variables
LexisNexis	Current Address, Name Updates, Contact Information
CLEAR	Current Address, Name Updates, Contact Information
National Death Index	Vital Status
Department of Motor Vehicles	Current Address, Name Updates, Contact Information
Accurint	Current Address, Name Updates, Contact Information
Social Security/Social Security Death Master File	Social Security verification, Vital Status
Vital Records	Names, Name updates, Dates of Birth, Dates of Death

Data System Mapping

When preparing to share data, you will need to understand what variables within the source systems are equivalent, and when they may need to be modified or standardized before data sharing. One easy example of this would be date formats. Dates can be formatted in a myriad of ways, but if we are planning to share and integrate data seamlessly between disparate systems, we should agree upon a standard format for those dates to prevent issues later on.

It is also important for the participants in a data-sharing program to understand the available data fields in the other system, and understand how those variables are relevant to their own data. Therefore, before sharing data it is important to have data stewards and data end-users gather to discuss their respective systems and come to a mutual understanding of the data fields that will be exchanged, and the intended use and value added of those data. Here we describe a potential

exercise that you can conduct in order to achieve an understanding of the scope and purpose of the data-sharing elements.

1. Create simplified data dictionaries, or a 'data element inventory'

If you have a data dictionary for your system, that is a great place to start. If you don't have a data dictionary, then this activity is a great way to build the first iteration of a data dictionary.! The end goal of this step is to have a simple inventory of all of the data elements in your system, in an easily digestible format that can be understood by someone who is not familiar with this system. See Table 4 for a template that you can use and adapt as relevant to your use-case.

Table 4. Sample Data Element Inventory

Data Field [system ID]	Data Field Short Descriptor	Variable Type	Details	Notes
What is the code for this data field within the system?	*What is a quick and easy descriptor for this element*	*Is this data field numeric, text (string), logical (boolean), etc.*	*A detailed description of the data field, as relevant*	*Any further information about this field that is important to note?*
First_Name	First name of client	String	First name as listed in client's primary profile table	First name 'alias' are available in a separate table, can be used to assist in matching algorithm
SSN	Social security number	String	Client's social security number as reported on case report form	These are most reliably collected for RWHAP clients, less reliable for STI cases - often not reported on labs and collected manually or through case report forms

2. Share your data element inventories, discuss, and dissect

Once you share your data element inventories with the other system owner, you can begin to discuss which variables will be important for your integration project. Each party can review the others' inventory, and make comments, questions, or highlight indicators of interest. This stage may require meeting regularly with the full team to clarify definitions and specifications. Parties will eventually be able to develop a wish list of indicators as well as understand the availability of person-level identifiers that can be used to link cases.

The end goal of this step of data mapping is to have a list per system of elements that will be included in the data share file, with a flag next to indicators that will be used as part of the linkage algorithm. A sample is included below in Table 5.

Table 5. Sample: Data Elements to Share

Data Source	Data Field (as coded in system)	Data Filed Short Descriptor	Notes
Enhanced HIV/AIDS Reporting System (eHARS)	first_name*	Client first name	Will consider adding alias names after the first match is done if we need to improve matching robustness
eHARS	SSN*	Social security number	Will send full and partial SSN
eHARS	Viral Load	Patient Viral Load result	Will share the most recent viral load for all clients.
eHARS	Viral Load date	Date associated with most recent viral load	

*Indicates data fields that will be utilized for data linkage

3. Document how you intend to use new information

Once you have a clear understanding of which variables you will be receiving as a result of the data match, it is important to ensure a clear understanding of how you will use those incoming data. The goal is to not share data just for the sake of sharing data but to ensure that all data sharing is intentional, and planned. Table 6 gives an example of how you may consider documenting the protocol for using newly shared data.

Table 6. Sample Data Processing Plan

Data Element (in other systems)	Associated Data Element in Source System	Processing Steps
viral_load	patient_viral_load	Compare associated viral load dates between the two systems 1. If the other system's viral load date is more recent, append this record to the patient file. 2. If the other system's viral load is the same as yours, no action. 3. If the other system's viral load date is older - query the patient file to see if this viral load is already stored in the system. If not, add this record. If yes, no action.

Data Sharing

When sharing data, the first distinction to make is whether this exchange is occurring internally (within an organizational unit) or externally (outside of your organization/agency). If external, then the parameters of your exchange should be clearly outlined within a data-sharing agreement. If internal, you may not need a data-sharing agreement, but there should still be an agreed-upon protocol for sharing information between systems.

For Internal Data Sharing

When sharing data internally, there will likely be less need for a strict file transfer protocol. However, before sharing data ensure that you are fully aware of any organization standards, requirements, etc.

For External Data Sharing

When sharing data externally, it is imperative that secure data-sharing protocols exist. Secure File Transfer Protocol (SFTP) allows for encrypted client-server communication via Secure Shell Protocol (SSH), which ensures a secure transfer of data between computers from outside the organization's network.
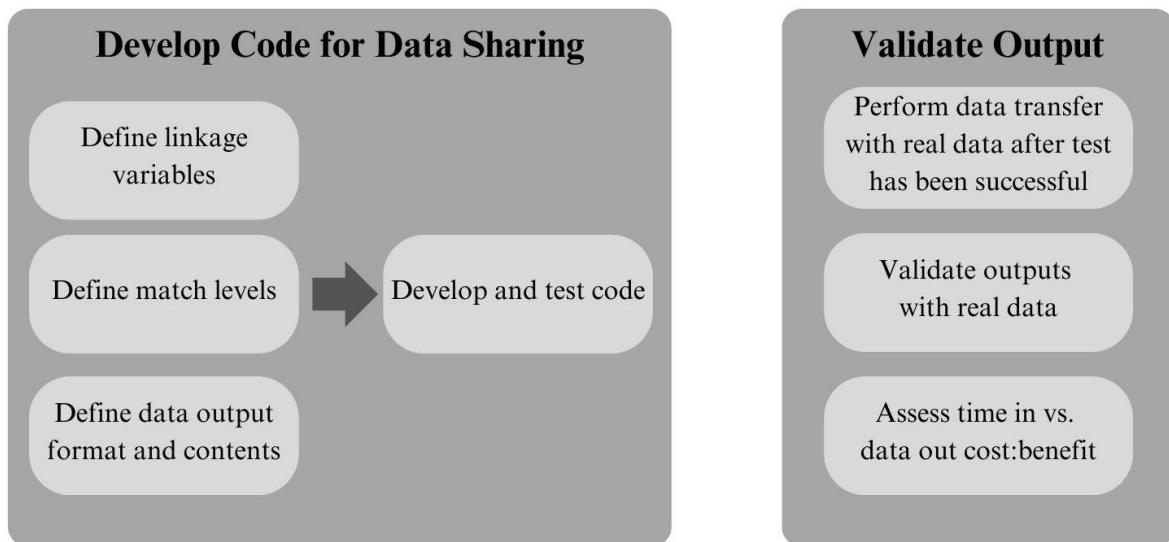
Data Linkage

Figure 2: An overview of data linkage processes with considerations for code development (left) and output validation (right).

HIV, STI, and RWHAP care activities are all conducted on similar populations, yet data collected for these populations are collected and stored by different systems and are not routinely shared or matched between organizations. Improved data linkage and synchronization of these data systems can provide greater accuracy and ability to assess HIV-related health outcomes, give a more complete and accurate representation of the HIV care continuum, and improve data-to-care activities. A brief overview of the data linkage process is shown above in Figure 2. and elaborated on below.

Develop Code for Data Sharing

When developing the code for data sharing, you must consider how complete the received variables are and how to standardize them. At times, though there are common variables that are collected in different databases, data may be collected in different forms. Examples of how to standardize variables include making sure that the variables that will be matched are either both character or numeric or if variables such as a person's name are in "all caps" or their DOB is formatted by year/month/day, that both datasets have them packaged the same way. It must also be considered that though surveillance data captures personal information that will identify a specific individual, there may be data entry errors that may cause mismatches or non-matches. To help identify or capture these errors, creating matching keys using the variables from the linkage datasets will allow a programmer to evaluate the level of the match, what data entry errors may have occurred, and what keys to include in the final matched dataset to make sure

everything will link correctly. Continue to develop and test the code until all linkage issues are resolved.

Validate Output

Once the linkage SAS code has been completely run, jurisdictions will decide what level of the match will be included in the final dataset. After the data match, address any incongruent data from the different source data systems. Use all data sources available to the jurisdiction to make final reconciliation decisions. Supplemental data sources can be used to verify or correct data. Consider which data sources are most highly trusted for different data elements. Create an optimal workflow for assessing new data linkages and determine which information you want to import from one system to another. Agree upon standards for which data sources are more trusted and where supplemental can be used to support existing DOH data. Identify when there is confidence in overriding data and when to not override data.

After the final dataset has been created, the matched dataset can be shared between HIV surveillance, STI, and the RWHAP personnel. The matched dataset should be validated with real data to confirm success and usability. This linked dataset can be routinely used to drive the creation or enhancement of program activities, as well as drive future research topics. As jurisdictions continue to identify future needs, advance how data is stored, and understand the cost-benefit of time dedicated to the match process versus the benefit received, jurisdictions will need to decide how often these activities will need to be conducted.

Data Storage

Routine data matches are great for updating data for reporting, but to continuously have access to integrated data it is important to identify an appropriate data storage solution. There are many considerations when choosing the design and architecture of your data storage solution. This section will walk through some of the decision points and will provide examples of benefits and limitations of various options. Ultimately, there is no single solution that will work for all data integration projects. The right solution will be tailored to each use-case and depends heavily on the specifics of the technological and human resources available.

Decision One: Integrated System versus Data Warehouse

When embarking on a data integration project, you may realize that you are constantly performing data matches on two separate systems that really would be more effective as a single integrated system. In these cases, it may be beneficial to transition from the two systems to a single integrated data system that can absorb the storage and user interface capabilities of both original systems. In other cases, you may find that you have lots of systems that work perfectly well on their own, and you just need a central 'window' into all of the systems in one space. In these cases, a data warehouse as a supplementary layer may be the best option.

Figure 3 provides a simple graphical depiction of the structural difference between the integrated systems and data warehousing options.
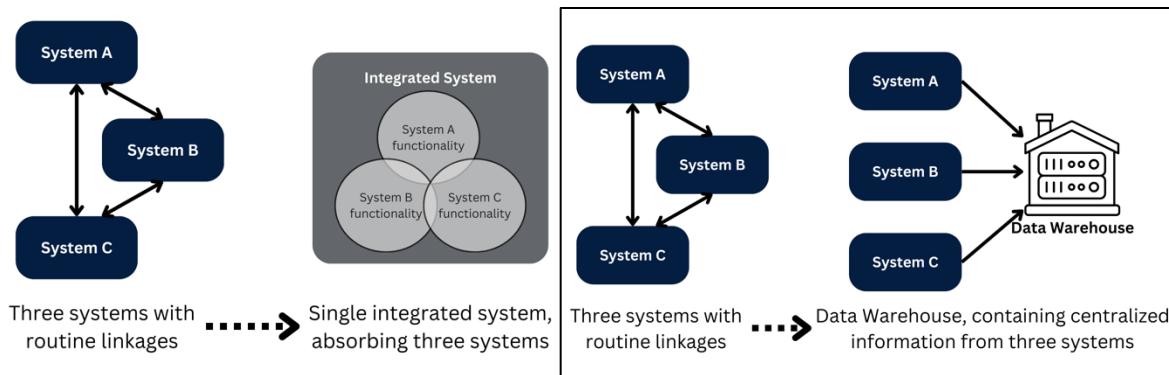
Figure 3. Integrated Data System (left) vs. Data Warehouse (right)

Decision Two: Home-Grown versus Off-the-Shelf

The home-grown approach allows for more customizability, but you have to have the in-house skill and capacity to do business analysis, requirements gathering, design, building, testing, and deployment. Off-the-shelf may be a bit less perfect, but then you are saving a lot of time and effort when starting off because the system is already designed. This way you just have to fill in the data and customize the layout. A middle ground would be a highly customizable off-the-shelf system, where you start with the template but you have high flexibility to make modifications and additions to make the product your own. This still requires highly skilled systems developers but will not require quite as much time in the design and build phases.

Decision Three: Internally Managed versus Software as a Service (SaaS)

You can host your system in-house, likely by your DOH's IT team or health systems team. You could also pay a SaaS vendor to host the system for you. Benefits of SaaS include having them manage your system maintenance, version control, patches, upgrades, etc. This way you don't have to rely on internal staff to do this, which is especially beneficial if you share your IT team with the entire DOH and are worried about response time for your system. However, depending on how your agency is organized, it may be beneficial to keep the system in-house. If your IT team is very strong and they are already managing other systems that you want to incorporate interoperability with, it may be easier to have them manage this one too so that they can be the ones to set it up within the existing architecture with the other systems.

Decision Four: Cloud Storage versus Local Server

Will your data live 'in the cloud' or will you host it on a local server? This probably is a decision for your DOH's information technology team; they likely have a preference and a standard for your office. More and more, agencies are electing to store information in the cloud. Previous concerns with cloud storage included security, but this is becoming less of a concern as the modern age of data is heavily cloud-supported. Hosting on a local server requires server maintenance, physical space, and if there are power outages you risk temporarily losing access to your data. This overview from the National Academy of Sciences discusses the measures that are taken to encrypt and protect cloud data.

Building your System

After you decide on the architecture of your integrated data storage solution, you will have to actually define the specifications needed. These specifications will be both technical and functional. For example: what is the volume of data that you expect to be in the system (technical), and how do the users interact with the system to view the data (functional). If you are looking for a fully integrated system, it is likely that you will spend a lot of time defining the necessary user interface, as the users will need to re-learn their current day-to-day activities with the original systems, within the new system. If you are pursuing a data warehouse (DWH), this is a supplementary tool that is designed to add value to existing systems, rather than replacing them altogether. For a DWH you will need to focus on what the added value of this system is, and how to design a user interface that will allow for this value to be realized.

Overall, the decision to pursue an integrated data storage solution will lead you down a road of business process analysis, documentation, requirements gathering, and even more documentation. The section below provides a framework of the types of in-depth assessments you will need to perform, but is not exhaustive.

Needs Gathering for your Integrated Data Storage Solution
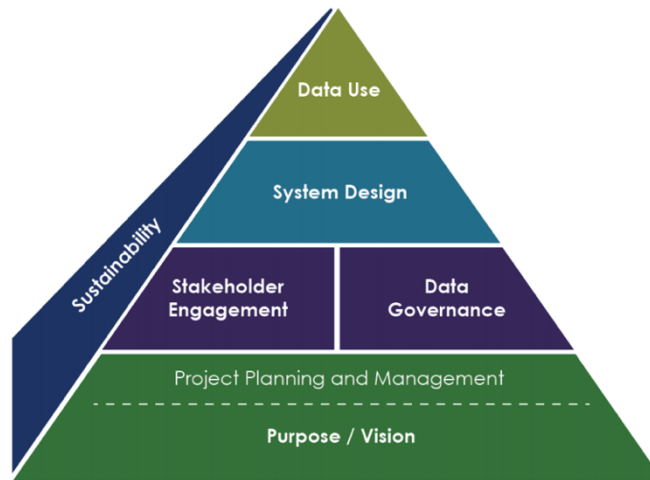
Action steps developed based on framework for data systems retrieved from:
https://nces.ed.gov/programs/slds/pdf/framework_for_data_systems_may2014.pdf

Technical Assistance Provider actions may include:

Work with jurisdiction to define the purpose/vision of data system <u>enhancement</u>:
● Define enhancement purpose – a clear vision, detailed goals, and key strategies.
● Define desired outcomes for the enhancement (data utility).
● Draft planning and management processes with purpose and vision of enhancement in mind.

Engage Stakeholders within the DOH
Bring all proposed enhancement documentation before stakeholders to:
- Receive feedback on the defined outcomes from stakeholders' perspective.
- Define purpose, roles, and responsibilities of stakeholders.
- Establish engagement and participation of stakeholders.
- Establish lines of communications between all stakeholders and related groups.
- Develop a communications plan for the enhancement, including a feedback loop.

Data Governance
- Discuss existing data governance within the group.
- If none exists, define a data governance process which includes:
  - Purpose
  - Staff roles and responsibilities – including decision making authority
  - An issue resolution/escalation process
  - A process for conducting meetings, including documentation
- Assess, define, document, and engage in in data quality processes and training.

System Design
- Implement project planning and management processes.
- Ensure appropriate access and privacy controls, potentially including:
  - Data-sharing agreements/memorandums of understanding data confidentiality policies
  - A role-based access model
  - A data security and monitoring plan
- Determine the software procurement process, if not already determined.
- System design processes:
  - Collection and storage
  - Data quality and data validation
  - Data warehousing
  - Data extract, transfer, and load (ETL)
  - Plan for data exchanges, data matching, data standards, and identity management

Data Use
- Create and analyze data usage via user metrics.
- Develop reports and documentation regarding altered processes, added system outputs, added user input, etc.
- Support users via system and data use training.
- Continue to collect feedback from users and stakeholders to guide further modifications to protocols, processes, etc.

Sustainability – ensure long-term usability of data system
- Promote widespread and consistent use of the enhancement/updated protocol.
- Assess "return on investment" of data system enhancement.

Data Governance Considerations

Data Governance
- Discuss existing data governance within the group(s) at the DOH.
  - If none exists, define a data governance process which includes:
    - Purpose
    - Staff roles and responsibilities – including decision-making authority
    - An issue resolution/escalation process
    - A process for conducting meetings, including documentation
- Assess, define, document, and engage in data quality processes and training.

Conclusion

Data integration helps collect data from multiple sources to gain better insights, increase communication, and improve health outcomes. The data integration process has impacts on data quality, data storage and security, and collaboration between different offices within departments of health. In HIV surveillance, reconciling disparate systems helps reveal useful data trends that can ultimately translate health data to impactful health program outcomes.

Implementation Guide 4: Best Practices for Data-to-Care

Introduction

Data to Care (D2C) is a public health strategy that uses HIV surveillance data and other data sources to identify a person with HIV who is not in care. The data is subsequently used to either link or re-engage that person into HIV care and any social services that they need to successfully move toward viral suppression and management of HIV through the HIV care continuum. The overall goal is to engage people with HIV in care and help those with HIV to reach viral suppression. People with HIV who are virally suppressed cannot sexually transmit HIV to their partners and can live longer and healthier lives. According to the CDC, viral suppression is defined as having less than 200 copies of HIV per milliliter of blood.[1]

Considerations for Data-to-Care

Data Sources

The main source of data as it pertains to this project will come from the enhanced HIV/AIDS Reporting System (eHARS). This data system is the primary software storage for all HIV-related data at the state level. At state health departments, there are many other potential data sources that can be utilized to provide the maximum amount of information needed for RWHAP recipients to effectively complete data to care activities. Some of these data sources are already matched to eHARS on a regular basis, in which case there may not be a need to do additional matches during the data exchange. Reviewing the frequency and timing of existing data system matches ensures that the RWHAP data exchange with eHARS adequately captures updates within the other data sources.

If the data source is not being matched to eHARS already (or is matched at an inadequate frequency), the HIV section must determine whether this data source can provide additional information that will benefit RWHAP D2C activities. Below are data sources that should be considered in developing data exchanges.

Potential Data Sources:
- AIDS Drug Assistance Program
- CAREWare Database
- Local Health Department HIV Surveillance
- Sexually Transmitted Infections (STI) Surveillance Data
- Medicaid Database
- Vital Statistics

RWHAP data matching with any of these sources would require data sharing agreements (DSAs) if they are not already in place.

---

[1] *HIV by Age: Viral Suppression*. (2022, July 1). Centers for Disease Control and Prevention. https://www.cdc.gov/hiv/group/age/viral-suppression.html

Providing Data

To begin, the RWHAP recipients will submit a list of people who have been identified as being a person with HIV in their data systems. A predetermined set of variables for each individual will be added to this list in order to provide a basis for matching efforts once the list is received at the state level. Example of predetermined variables are the following:

Data Sent by Part A (per client)
- Client ID
- Matching Variables – First Name, Last Name, DOB, SSN, Sex at Birth, Current Gender
- Phone number
- Current Residential address and date
- Most Recent Viral Load (VL)/CD4 result and date
- Vital status
- Out of Jurisdiction (OOJ) status

For data quality purposes, the recipients should add a date field to their data as much as possible, for example, "last known address of 5/2020". Both parties should define the most sustainable and confidential method for transmitting the sensitive data which contains personal identifiers, like establishing secure file transfer protocols (SFTP) with each recipient.

Receiving and Matching Data

Upon receiving the data, the Department of Health (DOH) will use the predetermined variables provided in the list to conduct a match to their eHARS data system. A validation of the matches will take place. Once the match is complete, a person should fall into one of the categories below:

- In Care
- Out of Care (OOC)
- Deceased
- Moved Out of Jurisdiction (OOJ)

The client ID and the number of persons in these categories will be sent back to the recipients. Additional information should be provided for those who are truly out of care, including any up-to-date data that was available in eHARS. The state health department must determine how to proceed with individuals on the recipient list that may not be in the eHARS database.

The DOH and recipients need to determine the minimum amount of data that will be needed for staff to conduct their linkage and re-engagement activities. Using the minimum amount of data will help to prevent security and confidentiality breaches. Additionally, special attention must be paid to the method for transmitting data back to the recipients to ensure that the mechanism is secure and sustainable.

Reviewing and Using the Data

Once the data has been matched with eHARS, it would be sent back to the recipients. Once the recipients receive the data, they will need to review it to see overall numbers and what cases need follow-up. Depending on the number of cases provided, it may not be feasible for the recipient staff to follow up on everyone. This may be due to staffing shortages or an unreasonable volume of OOC clients to follow up in a designated time period. Therefore, recipients may want to use the data to create a linkage-to-care prioritization list. This list would be compiled based on priority categories that the recipient team determines. Examples of priorities can be but are not limited to:

- Time Frame
- Risk
- Geographical Location
- Viral Load Count

Using priority groups will help to ensure that linkage staff gets maximum utility from the data they have been provided. Maximizing the effort invested in linking and/or re-engaging patients may lead to better outcomes. It can be helpful to center the prioritization around special populations that are largely impacted by HIV in the jurisdiction, who may not be reached through other channels.

These data matches to ensure that the linkage staff has the most updated and reliable information. This will allow them to be more effective and efficient when beginning their field investigation process of linking these individuals back into HIV care. Timing will determine how useful the returned data list is at the time. For example, a list run six months ago may not be useful to use, as HIV information can change frequently and now may be outdated. Therefore, manual reviews may need to be done on each individual on the list in order to have more accurate and up-to-date information at the time of initiation. In order for these manual reviews to be done, the RWHAP recipients may need access to searching mechanisms such as:

- LexisNexis (Accurint)
- Department of Motor Vehicles
- Unified hospital systems
- Other data systems

All information should be documented to ensure all the steps of the process are in one place. Documentation can help when providing feedback data to HIV Surveillance pending the outcome of the investigation.

Conclusion

D2C activities help public health professionals link people with HIV to care while using available time and resources as effectively as possible. This helps improve care outcomes by finding gaps in surveillance to reach those more likely to fall out of care. Data linkage activities bolster shared surveillance and communication among states and RWHAP recipients, aiding D2C activities, and improving health outcomes for people with HIV.

Implementation Guide 5: SAS Linkage Code Tool

Surveillance data has long been used to describe demographics and outcomes along the HIV continuum of care among people with HIV but it is known that data is not only collected for people with HIV in this entity. Analyzing HIV surveillance has been a valuable tool for jurisdictions to assess trends, but the sole use of this data source to evaluate successes and gaps may not tell the full story of health outcomes among those with HIV. The synchronization of data systems will lead to:

- Provide greater accuracy and ability to assess HIV-related health outcomes among people with HIV.
- Give a more complete and accurate representation of the HIV care continuum.
- Improves data-to-care activities.

Preparing Data Files

One of the first steps in conducting the data linkage process between different data sources is to prepare the data files to be used. Before files have been pulled from the data collection system, jurisdictions must decide what key variables will be used to perform the match. Some key variables that are routinely used for a linkage include first and last name, DOB, and sex at birth. Social Security number (SSN) is also used, depending on how routinely and completely this variable is reported in the data collection system. Jurisdictions are able to include other variables that both data sources capture to improve the match results, if necessary. After variables have been finalized, staff extract data from the system and transfer the file to the personnel who will be performing the data linkage activities using a secure data transferring system.

Deciding What to Include in the Linkage SAS Code

There are several things to consider when writing a SAS code to complete a data linkage activity. First, the programmer may consider how complete are the variables received; adding syntax to the first check to see if there are variables received that have responses that are not left blank will be helpful. Secondly, the programmer must consider standardizing data variables; making sure that the variables that will be matched are either both character or numeric, or if variables such as a person's name are in "all caps" or their DOB is formatted by year/month/day, that both datasets have them packaged the same way. Finally, it must be considered that though surveillance data captures personal information that will identify a specific individual, there may be data entry errors that may cause mismatches or non-matches. To help identify or capture these errors, creating matching keys using the variables from the linkage datasets will allow the programmer to evaluate the level of the match, what data entry errors may have occurred, and what keys to include in the final matched dataset.

Standardizing Data Variables

After deciding what to include in the linkage SAS code and having all datasets pulled and ready, the programmer will import the datasets into SAS. But before the match can take place, the programmer must first standardize the variables to be matched to make sure that they will link correctly. At times, though there are common variables that are collected in different databases, data may be collected in different forms. As previously stated, variables may be collected as a

numeric or a character, stored in the system in all caps, or have dashes, so the programmer must make sure that these measures are formatted across all variables used so that the variables can successfully match. Some formatting that may need to occur include:

- Upper case all character variables
- Remove spaces and special characters from name variables
- Separate DOB into day, month, year
- Compress DOB
- Standardizing the length of variables

Below are examples of SAS syntax in standardizing variables:

To upper case names
data ehars_data; set ehars_data; l_name = upcase(LAST_NAME);
      f_name =upcase(FIRST_NAME); drop last_name first_name; run;

To remove dashes from names
data ehars_data; set ehars_data;
l_name2=compress(l_name, "-"); f_name2= compress(f_name, "-");

To standardize length of variables
data ehars_person; set ehars_person;
leng = length(ssn);
if leng= 11 then do;

ssn2= compress(ssn,"-"); end; run;

To compress DOB
set ehars_person;
        BDMon=Substr(DOB,5,2);
        BDDay=Substr(DOB,7,2);
        BDYr=Substr(DOB,1,4);
        Bday=compress(BDMon||BDDay||BDYr);

Please note that although what is presented above is formatting variables in the "eHARS data" dataset, these same formats must be completed for all other datasets that eHARS data will be matched.

Creating Variable Matching Keys

To specifically identify how HIV cases were matched between the two data sources, "keys" may be used to consider how data can be matched and consider what will be included in the final dataset. Taking into consideration that there may be data entry errors, some keys may be created to capture variations in first name, last name, or DOB

Creating and formatting variables for DOB
Data Trunc_eHARS;

```
        length _all_ $20.
                BDMon $2.
                BDDay $2.
                BDYr $4.
                BDay $8.;
        set ehars_person; BDMon=Substr(DOB,5,2); BDDay=Substr(DOB,7,2);
                BDYr=Substr(DOB,1,4); Bday=compress(BDMon||BDDay||BDYr);
```

Creating error variables for first name
```
if first_name ^=" " then do;
                Fnme6=Substr(first_name,1,6);
                Fnme1=Substr(first_name,1,1);
                Fnme2=Substr(first_name,1,2);
                Fnme2_8=Substr (first_name,2,7);
                Fnme2_4=Substr(first_name,2,3);
                Fnme3=Substr(first_name,1,3);
                Fnme4=Substr(first_name,1,4);
                fnme4_8=Substr(first_name,4,5); end;
```

Creating error variables for last name
```
if last_name ^=" " then do;
                Lnme1=Substr(last_name,1,1);
                Lnme2=Substr(last_name,1,2);
                Lnme3=Substr(last_name,1,3);
                Lnme4=Substr(last_name,1,4);
                Lnme3_8=Substr(last_name,3,6);
                Lnme4_8=Substr(last_name,4,5);
                Lnme2_4=Substr(last_name,2,3);
```

Creating keys
```
Key1=Compress(last_name||First_name||bday||sex||ssn)*;
Key2=Compress(last_name||Fnme6||bday||sex);
Key3=Compress(Lnme1||Lnme3_8||Fnme2_8||bday||sex);
Key4=Compress(Lnme1||Lnme3_8||Fnme2_8||BDMon||BDYr||sex);
Key5=Compress(Lnme1||Lnme3_8||Fnme2_8||BDDay||BDYr||sex);
Key6=Compress(last_name||Fnme3||bday);
Key7=Compress(first_name||last_name||bday||sex)
key8= ssn2; *
Key9=Compress(Lnme3||Fnme3||bday);
```
*Included if jurisdictions collects complete SSN

Though the "keys" created above are examples how jurisdiction can identify ways that a case
will match, this is not an exhaustive list and more keys may be added to suit other possible
matches.
Full SAS code can be found **here**

Completing the Match and Finalizing the Matched Dataset

After completely running the SAS code, the data will be output by the different keys created. Once the linkage SAS code has been completely run, jurisdictions will decide what level of match will be included in the final dataset. Based on the keys that were created above, jurisdictions will decide only what are considered "exact matches" or matches that only include full first name, full last name, full DOB, and sex at birth (include SSN if used). Some jurisdictions may include exact matches in the final dataset but also include keys that also match that had some possible errors in the first name, the last name, or the DOB. It is of note that one way to decide whether to include other keys is to evaluate the "non-exact" match dataset by first comparing the number of cases that were included in the output datasets for each key. If the majority of the matched cases were in the datasets that are considered "exact", eliminating the other non-exact may be easier, but if there are more matched cases in the non-exact datasets, then first checking the matching keys syntax should be reevaluated to be sure that there was no misspelling were included and if no misspellings were included, decide at what level of non-matches are to be included.

Moving Forward

After the final dataset has been created, the SAS programmer is able to share the matched dataset between HIV surveillance, STI surveillance, and the RWHAP personnel. This linked dataset can be routinely used to drive the creation or enhancement of program activities, as well as drive future research topics. As jurisdictions continue to identify future needs or advance how data is stored, jurisdictions will need to decide how often these activities will need to be conducted.

Evaluation Guides

Evaluation Guide 1: Data-to-Care Outcomes Monitoring Spreadsheet

As part of the Georgetown University HIV/STI Data Integration for Improved Ryan White HIV/AIDS Program Outcomes project, several states introduced a data sharing pathway between the central surveillance entity and local level case management entities. The purpose of this data sharing is to ensure that local entities have timely, comprehensive data about the care engagement status of their clients, in order to better prioritize re-engagements with people with HIV who are truly out of care. The purpose of this monitoring guide is to provide a template for similar data-sharing projects to assess the effectiveness of the data sharing in identifying accurate care status, and for those confirmed out of care post-match, what is the outcome of re-engagement activities?

Overview of Spreadsheet Tool:

Sheet One: Context
Sheet One can be filled out before the match occurs, by the local entity (i.e. RWHAP Part A) and describes the data that are being submitted for the match. Column C details the inclusion criteria of the clients being sent to the state for a match. Columns D and E detail how many of the clients are suspected to be out of care (OOC) versus active before the match with the state is performed.

Sheet Two: Match Results
Sheet Two can be filled out by the state surveillance team after conducting the match with the local entity's client list. This sheet will describe the results of the match, in aggregate format. Columns B and C detail the total number of cases sent by the local entity, and how many of those cases match with the state's eHARS database. Column E lists how many of the clients from the local entity are deceased according to eHARS data. Columns F through H detail the updated care status of the remaining clients based on the state's data.

Sheet Three: Outreach Outcomes
Sheet Three can be filled out by the local entity once the state returns the updated data file, and the local entity has had time to conduct re-engagement efforts with the confirmed out of care clients. Columns C and D give information about individuals that were not located (total, and with a field visit). Column E totals the number of clients who are found to be dead after case manager investigation. Columns F through H give information about clients that are located to be out of jurisdiction, with unknown care status. Columns I through O gives information about individuals who were located as already being in care, with several breakdowns to provide insight into why the state may not have been aware of this status. Finally, Columns P through U give information about the outcome for clients who were identified as not in care.

[Link to Data-to-Care Outcomes Monitoring Spreadsheet](#)

Conclusion:

This Data to Care Outcomes Monitoring Spreadsheet tool can be utilized to understand the data match-to-linkage to care pipeline. The general logic of this tool can be modified as needed to accommodate various states' needs. Utilizing shared data to bring more people into care is an important part of reducing unnecessarily wasted efforts, including duplicative linkage efforts of those in care elsewhere, ultimately saving time and more likely to have resources allocated to

those truly OOC.

Evaluation Guide 2: Internal Evaluation

Why Evaluate?

Project evaluation is one of the key pillars of project implementation itself. Evaluation can function alongside implementation, at the close of your project's period, or both. These crucial steps allow you to assess the quality and effectiveness of your efforts to implement the project, identify areas of needed improvement, and update your goals moving forward. This guide will provide a brief overview of the components of a successful evaluation.

Three Types of Evaluation

There are three types of evaluation. You can select just one or use a combination of the three. These are Strategic Evaluations, Planned Evaluations, and Spontaneous Evaluations.

A. A **Strategic Evaluation** is an in-depth assessment answering the question: "To what extent is our project reaching its desired goals?" This evaluation is needed to assess progress and redirect goals and objectives as necessary according to evaluation findings.

B. A **Planned Evaluation** is a standard evaluation tool used to monitor the project as it goes. These evaluations vary in depth and scope but typically provide feedback on a regular basis (i.e., quarterly or annually) to inform the project's progress and future goals.

C. A **Spontaneous Evaluation** is used to assess issues or unforeseen events that arise throughout implementation that need further investigation. This evaluation can help explain the *why* and *how* of a particular issue that may influence implementation outcomes. For example, a spontaneous evaluation could have been used at the start of the COVID-19 pandemic to assess the expected impact of the pandemic on project implementation.

Evaluation Guide

1. **Set objectives**: Depending on the stage of your project, you might want to ask different evaluation questions. Before you conduct any evaluation, you must first set the objectives you are looking to achieve in the evaluation. This could be a list of questions or categories you're hoping to use the evaluation to better understand.

2. **Develop methodology**: There are a number of tools that can be used for project evaluation, each depending on the overarching objectives of the evaluation and the questions being asked. Some examples are surveys, Key Informant Interviews (KIIs), or Focus Groups.

3. **Select survey population and deploy tools**: Identify the group or groups of participants, stakeholders, or individuals involved in any area of the project you are hoping to evaluate. Ensure that the appropriate people are contacted in the appropriate topic areas.

4. **Analyze results**: The analysis will need to be structured and tailored around the type of data collected, qualitative, quantitative, or both.
5. **Write up results**: Once the analysis is complete, an evaluation report is necessary to consolidate the findings of the evaluation and analysis. The results should detail the challenges, successes, and gaps identified in project implementation.
6. **Develop recommendations**: Based on these identified challenges, successes, and gaps, recommendations should be developed on how to best improve the effectiveness of the project.

Appendix

Appendix 1: COP Questionnaire Guideline

Communities of Practice (CoP) Guiding Questions
Considerations for creating and maintaining CoP

Define CoP

- What are your reasons for forming a CoP? What are the issues that you would like to address via a CoP?
- Who should be involved in the CoP? How does the team define CoP? What are the ways to structure CoP?
- What are the essential characteristics of a successful CoP?
- What strategic communication skills are needed to implement effective CoP?

Address Challenges

- How do we address the challenges with maintaining productive CoP due to competing priorities or staffing constraints?
- What are ways to resolve challenges collaboratively?
- How have public health emergencies (i.e., COVID-19) impacted staffing in the department, and can we learn collaboratively?

Sustain Impact

- What makes CoP effective? What teamwork needs to be developed and implemented?
- How do we monitor impact of CoP? What is the intended outcome of the group?
- Are we prepared to facilitate routine, active communications with expert advisors, peers, and other stakeholders?

Appendix 2: Data Sharing Agreement Template

[State/Department of Health name]
And
[Provider/Facility name]
HIV Surveillance Data Sharing Agreement

This is a Data Sharing Agreement between [*enter Provider/Facility name*], herein referred to as the "Recipient" and [*enter State/Department of Health name*], herein referred to as "State". When referred to collectively, the State and the Recipient are referred to as the "Parties."

1. **Purpose of Agreement:**
    The objective of this Agreement is to outline the roles and responsibilities of the Parties in the project linking Recipient data with the data from the State.

    The State ascertains HIV case data by receiving reports of potential cases from clinical providers, laboratories, and other public health providers and then collects basic demographic and risk information on these cases.

    The Recipient is an HIV service provider, responsible for ensuring that individuals in its care receive the necessary care and treatment, both to promote the health of the clients and limit the opportunity for viral transmission.

    Data accuracy and synchronization have been documented to lead to better care outcome activities, such as increased linkage to care, increased engagement and re-engagement in care, and ultimately, increased viral suppression among people with HIV. Enhanced synchronization of disparate data systems assists in confirming residence and care patterns among persons diagnosed with HIV, including persons lost to follow-up and/or in care in other jurisdictions.

    The State offers routine opportunities for HIV service providers to submit their HIV patient data to enhance the data utilized to prioritize and conduct follow-up with clients who are not actively receiving care/treatment.

    Improved data linkage and synchronization will lead to more data accuracy and will ultimately inform a timely review of client-level care status. The synchronization of STI and HIV surveillance data will be used to improve the capacity of health departments and Ryan White HIV/AIDS Program (RWHAP) clinics to prioritize resources for linking and re-engaging people with HIV into care.

2. **Legal and Contractual Authority:**
    a. *[Cite laws and codes requiring reporting of certain diseases, statutes, etc. for applicable state]*

3. **Understanding of the Parties:**

a. The Recipient will submit its data file in the format defined in **Appendix A**, via a secure data transfer program that is approved by both Parties. Monthly deadlines for file submissions will be clearly defined by the State. Files will be extracted from the secure folders by State staff on the monthly deadline date. **Any files submitted after the deadline will not be processed until the following month's deadline.**

b. The designated technical staff from the State will perform the linkage on a routine basis according to the State's confidentiality and security policy [*Define/attach accordingly – Attachment 1*].

c. The State will conduct data linkages no more frequently than on a monthly basis [*or timeline otherwise agreed upon*]. The Recipient will be given a window for data submission each month, after which new data submissions will not be linked until the following month. The Recipient may elect to submit data for linkage at a frequency determined independently by the Recipient, in alignment with the monthly upload dates provided by the State.

d. A linkage algorithm including probabilistic and deterministic linkage strategies will be used. The linkage includes alternative spellings and social security numbers to improve the likelihood of a match. (**Appendix B** lists the variables used for the linkage and matching algorithm. [*Adjust as necessary for each agreement per state/program goals.*])

e. After the linkage and reviews are completed, the following action will occur:
   i. Using the same secure data transfer program as described in Paragraph 3.a, a line list of matched outcome data will be sent back to the Recipient, which will provide an updated disposition for each of the clients sent by the Recipient. **(Appendix C)**. The matched outcome data will be returned to the Recipient only in the event that a file is submitted by the Recipient by the monthly deadline, as described in Paragraph 3a. Therefore, the maximum frequency of this file exchange is monthly – and is dependent on initial submission by the Recipient, as described in Paragraph 3a.

f. Both Parties will protect the data at all times in accordance with the laws and rules cited in Paragraph 1 above. The data shall not be disclosed, released, revealed, shown, sold, rented, leased, or loaned to anyone outside of this agreement.

4. **Unauthorized Disclosure**:
   a. Only authorized personnel will have access to the data exchanged under this agreement. These personnel will be advised of the confidential nature of the information and the civil and criminal sanctions for noncompliance contained in the applicable federal and State laws.
   b. The matched records will be processed in a manner such that unauthorized persons cannot access the records by means of a computer, remote terminal, or other means.
   c. Information exchanged between the parties shall be used solely as permitted by the laws and rules cited in Paragraph 1 above.
   d. No files provided by the Parties will be used to extract information concerning individuals for any purpose not specified in this Agreement. These files will not

be duplicated or disseminated without the written consent of the Parties that furnished the data.

e. Data protected by the laws and rules cited in Paragraph 1 above are being exchanged under this Agreement. Any usage or dissemination outside of the terms of this Agreement is prohibited by law. In the event that either the State or the Recipient determines that the other Party has made an unauthorized disclosure of the data provided, the opposite party may request a formal response to the allegation of unauthorized disclosure, a corrective action plan formulated to alleviate the possibility of any future unauthorized disclosure, and removal of the data from the system. It may also prohibit further release of data to the party.

f. The Parties acknowledge that criminal penalties may apply if it is determined that either party or any individual employed or affiliated therewith, knowingly and willfully obtained the file(s) or data under false pretenses.

g. If a Party's data is subject to a security breach, it will notify the other Party immediately.

5. **Personnel**: Each Party is responsible for all costs of its personnel, including pay and benefits, support, and travel. Each Party is responsible for the supervision and management of its personnel.

6. **General Provisions**:
   a. **Points of Contact**: Promptly upon the execution of this Agreement, both Parties shall define points of contact that will be the primary and secondary persons responsible for coordinating its implementation. Each Party will communicate, in writing, specific contact persons as well as changes in point of contact to the other Party. Contact information shall include a physical address, mailing address (if different), email address, and telephone number.
   b. **Correspondence**: All correspondence to be sent and notices to be given pursuant to this Agreement will be addressed to the points of contact defined pursuant to Paragraph 6.a and to:

      i. If to Recipient: *Contact and Address OF RECIPIENT*

         _____

         _____

      ii. If to the State: *Contact and Address OF STATE*

         _____

         _____

   c. **Funds and Manpower**: This Agreement does not document or provide for the exchange of funds or manpower between the Parties nor does it make any commitment of funds or resources.
   d. **Modification of Agreement**: This Agreement may be modified only by the written agreement of the Parties, duly signed by their authorized representatives. This Agreement may be reviewed routinely as needed.

e. **Termination of Agreement**: This Agreement may be terminated in writing at will by either Party.
f. **Entire agreement**: It is expressly understood and agreed that this Agreement embodies the entire understanding between the Parties regarding the subject matter of the Agreement.
g. **Effective date**: The Agreement takes effect beginning on the day after the last Party signs.
h. **Expiration date**: This Agreement shall remain in effect *sine die* from the date of execution unless one Party requests a change or terminates the Agreement.

Review: The signature below signifies that the content of this Agreement and any document(s) that accompany it have been reviewed and approved by the State and the Recipient.

Signature of **Recipient** Approval

_____   _____
Name, Title   Date

Signature of **State** Approval

_____   _____
Name, Title   Date

Appendix A: Data Submission File Data Fields
[to be adjusted as needed per program needs]

Data Definitions (Recipient to State) [example]

Variable	Format and Explanation
ClientID*	Clinic specific identifier, State does not retain this field but returns it to the Recipient at the conclusion of the match
First_Name*	Client first name, legal preferred
Last_Name*	Client last name, legal preferred
DOB*	Client date of Birth, in MMDDYYYY format, no spaces, slashes or dashes
SSN†	Client social security number, no spaces, slashes or dashes

*Required variables. If these data fields are not sent, the algorithm will not match the associated record to STATE data.
†Persons reported without SSN may be less likely to find a match to STATE data. Partial SSN will be accepted only if submitted as last four digits of SSN without any leading placeholders for missing values (ex: 1234 instead of XXX-XX-1234 or 000001234). Clients without SSNs should be submitted with a blank field.

Data Submission Excel Format (Recipient to STATE) [example]

ClientID	First_Name	Last_Name	DOB	SSN†

Appendix B: Variables to be used for the linkages *[example, adjust per program needs]*
1. Last Name
2. First Name
3. Date of Birth (DOB)
4. Social Security Number (SSN)

Appendix C: Data Feedback File Data Fields *[example, adjust per program needs]*
Data Definitions (State to Recipient)

Variable	Format and Explanation
Disposition	0 - No Recorded CD4 count or Viral Load within last six months
	1 - Has had a CD4 count or Viral Load within the last 6 months, at a location besides your clinic
	2 - Has had a CD4 count or Viral Load within the last 6 months, at your clinic
	3 - Dead
	4 - Moved Out of State
	5 - Not Found in STATE Systems
Date of Death	Date the client was deceased, **only** for clients with disposition 3
Last Viral Load, Virally Suppressed	(Yes/No) At the clients last viral load, they were virally suppressed (<200), **only** for clients with disposition 1 & 2

Data Submission Excel Format (State to Recipient)

Client ID*	First_Name*	Last_Name*	DOB*	SSN*	Disposition**	Date of Death**	Last Lab Virally Suppressed**

* Data supplied from Recipient
** Data returned from STATE

Attachment 1: State Confidentiality and Security Policy
[Attach any state security protections for servers and any confidentiality and security guidelines here.]

Appendix 3: Standard Operating Procedure Template

*Data Match Process between eHARS and STI Data Systems**

just an example, not an active protocol

SOP Background:
[Provide relevant information on purpose, definitions, and scope of the SOP.]

Purpose

[This section should outline the purpose for data sharing including relevant details on current practices within the state HIV/STI surveillance systems.]

This standard operating procedure describes the process by which the [agency] links HIV and STI data. At the state level, HIV data is housed in the Enhanced HIV/AIDS Reporting System (eHARS) separate from other STI data. Data in disparate systems may mean that one system has missing information that could otherwise be available. This protocol defines a monthly data exchange process that will enable the HIV and STI data managers to update their respective data systems, as well as gain an understanding of comorbidities among people with HIV in the state.

Definitions of Data Systems and Terms

[Repository for definitions and terms for this document to provide context for any jurisdiction-specific data systems and technical definitions.]

Include any definitions, context, and background.

eHARS. The Enhanced HIV/AIDS Reporting System (eHARS) is a browser-based, CDC-developed application. It is a document-based data collection and management system that assists health departments with the management of the state's HIV/AIDS cases. Through this system, jurisdictions can run reports and analyses, as well as facilitate the transfer of data to the CDC.

**STI Surveillance System.** Define the jurisdiction's STI surveillance system and how it is stored and can be integrated with eHARS.

SAS. Analytics software

SOP Overview:
[Detail data flows through acquisition, liking, validation, and use within the jurisdiction.]
The following graphic shows the thematic process of HIV and STI surveillance data acquisition, linkage, validation, and use.
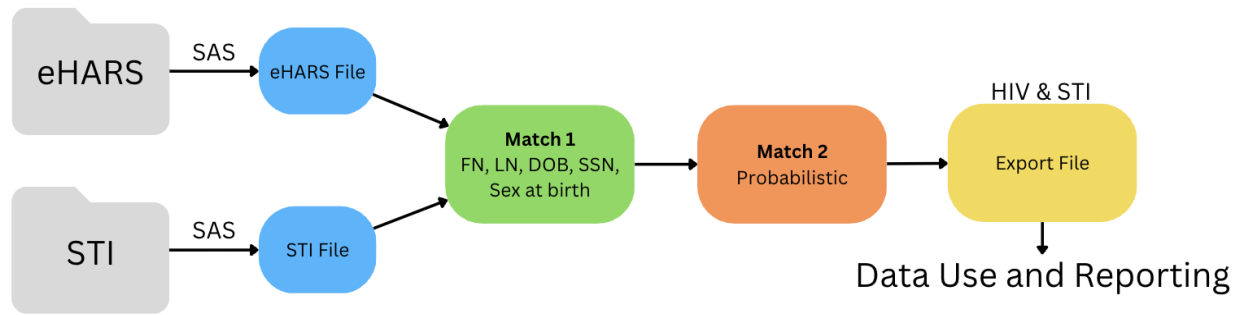
Data Processes
[Define terms of data sharing between entities. Include how data will flow, how often, and when and how data will be returned. See a template for a Data Sharing Agreement on page 69.]

Data Sharing Agreement

[Reference any legal agreements that were needed to enable this data sharing process.]

This SOP does not require a data-sharing agreement, as both parties are housed within the umbrella agency of communicable diseases. This SOP serves as a written agreement to the data exchange processes outlined in the document.

Define Data Variables

[Define data variables determined for matching between data sources in this jurisdiction. You should also include reference to data dictionaries of the source systems so that each party is aware of the contents of the export file (not depicted in this SOP).]

The variables in below will be used for matching between the two systems.

Example:

Variable	Format and Explanation
SSN†	Client social security number
First_Name*	Client first legal name
Last_Name*	Client last legal name
DOB*	Client date of Birth

Data Extraction from Source Systems
[Reference code to run data extraction e.g., where SAS code to run data extraction is located, how to run the code, and where the resultant file will be saved.]

eHARS extract: For the eHARS extract for this data sharing process, use the SAS code titled 'eHARS-Data-Export-for-HIV-STI-Integration', located on the shared drive pathway: [insert pathway here]. Once a month, the HIV data manager will use this SAS code to extract an auto

formatted data set from the most recent eHARS export. Each month, the user must change the global variable for filename and date to correctly direct the SAS code to pick up the most recent eHARS data export. Once the SAS code is complete, the exported data table will be available within the shared drive at [insert file path]_eHARS-monthly-export_MM-DD-YYYY.csv.

STI system extract: For the STI system extract for this data sharing process, use the SAS code titled 'STI-Data-Export-for-HIV-STI-Integration', located on the shared drive pathway: [insert pathway here]. Once a month, the HIV data manager will use this SAS code to extract an auto formatted data set from the most recent STI system export. Each month, the user must change the global variable for filename and date to correctly direct the SAS code to pick up the most recent STI data export. Once the SAS code is complete, the exported data table will be available within the shared drive at [insert file path]_STI-monthly-export_MM-DD-YYYY.csv.

Data Linkage
[If applicable, define matching protocol with variables. Detail step-by-step how HIV and STI data are matched and come through the system for further checks.]

Data will initially be linked via SAS code [file path here], which will look for matches based on first name, last name, date of birth, full SSN, and sex at birth. These cases will be designated exact matches, and exported in a file located at [insert file path]_exact-matches_MM-DD-YYYY.csv.

A secondary probabilistic match will be done with the remaining unmatched cases, using LinkPLUS. All matches over a 70% match will be accepted as a true match. Matches scoring between 50% - 69% will be manually reviewed (depending on the volume). Matches scoring below 50% will be discarded.

Data Use
[Describe how you expect data that is shared in this process to be utilized by your counterparts. If you have restrictions on how they can use the data, be sure to include those specifications in this document.]

Linked HIV-STI data will be used monthly by each team to make updates to their source systems as they see fit. Annually, the linked data will be used to create a comorbidity report that details the incidence of Chlamydia, Gonorrhea, and Syphilis among persons with HIV.

Challenges
*[Document challenges such as unknown data quality, lags, and missing data that may have
hindered any HIV-STI data linking efforts.]*

Each department should be conducting their own independent rigorous quality assurance activities to ensure that they are providing their counterparts with the highest quality data

attainable. However, each party is responsible for responsible data use - and should conduct quality assurance on the matched data prior to updating their source data.

Enforcement and Revision Plans

[Describe your expectations for how this SOP will be enforced, and how frequently the SOP should be reviewed by both parties to make any needed adjustments or refinements]

This SOP is a written agreement between the HIV and STD teams to adhere to the processes outlined in this document. If one team is unable to perform the agreed upon activities in a given month, written notice must be given at least 1 week prior to the expected data extraction date. This SOP will be reviewed annually by both parties, at which point recommendations for changes or revisions will be communicated, discussed, and enforced as agreed upon.

Appendix 4: Work Plan Template

Work Plan *for*
[enter project title/work group name]

Gantt Chart Timeline

Proposed Time Allocation: 3 Months

Task	Responsible Party	Month 1				Month 2				Month 3			
		1	2	3	4	1	2	3	4	1	2	3	4
Task Area 1													
Task 1.1		▉											
Task 1.2			▉										
Task 1.3				▉									
Task Area 2													
Task 2.1		▉	▉	▉									
Task 2.2					▉								
Task Area 3													
Task 3.1					▉								
Task 3.2						▉							
Task 3.3							▉	▉					
Task 3.4								▉					
Task Area 4													
Task 4.1									▉	▉			
Task 4.2										▉			
Task 4.3											▉		
Task Area 5													
Task 5.1										▉	▉		
Task 5.2											▉		
Task 5.3													▉

GEORGETOWN UNIVERSITY | Center for Global Health Practice and Impact

Context and Overview of Proposed Activities

Task Area 1
[Context of this task area]

Proposed Activities

[Any further information about the tasks listed under Task Area 1 that are important - such as who is responsible for carrying out the task, any technical information, links to relevant documents, etc.]

Task Area 2
[Context of this task area]

Proposed Activities

[Any further information about the tasks listed under Task Area 2 that are important - such as who is responsible for carrying out the task, any technical information, links to relevant documents, etc.]

Task Area 3
[Context of this task area]

Proposed Activities

[Any further information about the tasks listed under Task Area 3 that are important - such as who is responsible for carrying out the task, any technical information, links to relevant documents, etc.]

Task Area 4
[Context of this task area]

Proposed Activities

[Any further information about the tasks listed under Task Area 4 that are important - such as who is responsible for carrying out the task, any technical information, links to relevant documents, etc.]

Task Area 5
[Context of this task area]

Proposed Activities

[Any further information about the tasks listed under Task Area 5 that are important - such as who is responsible for carrying out the task, any technical information, links to relevant documents, etc.]

Appendix 5: Evaluation Survey Template

The HRSA SPNS HIV/STI Data Linkage project (HRSA-19-039) consisted of both an internal and external evaluation. Each evaluation was conducted through a series of interviews, surveys, and aggregate data analysis.

The goal of this evaluation survey is to assess the impact of technical assistance provided to jurisdictions partaking in the project. This is part of an internal evaluation conducted by the technical assistance provider and is essential to understanding the impact of the project as a whole and helps to inform improvements, recommendations, and next steps.

Survey

I understand the information that I have read and by taking this survey, I voluntarily consent to participate in this study. (Please check the box below if you consent to participate in this survey)

I consent
I do not consent

Effectiveness and Use of Communities of Practice

Did you participate in any of the Communities of Practice?

Yes
No

Which of the following Communities of Practice did you participate in? *Select all that apply*

Challenges to Program Implementation
Data Integration
Business Process Development
Data to Care

To what extent have the Communities of Practice made your daily work easier? Please explain.

Challenges to Program Implementation

Data Integration

Business Process Development

```

```

Data to Care

```

```


Do you think you will continue to hold Community of Practice meetings after the project period ends?

Yes
No

If no, why?

```

```


Would you recommend participation in Communities of Practice to other jurisdictions? *Select all you recommend.*

Challenges to Program Implementation
Data Integration
Business Process Development
Data to Care
None of the above

If none, please explain why.

```

```

Process for TA Provision

Was the needs assessment conducted at the beginning of the project helpful in defining the support you needed from the technical assistance team?

Yes
No

In what ways was the needs assessment helpful in defining the support you needed from the technical assistance team? *Select all that apply*

It was clear that could be provided
The needs assessment questions were relevant to our jurisdiction
It was helpful in identifying the support we needed
Other, please specify:

```

```

If no, why not? *Select all that apply*

It was not clear what could be provided
We were not sure what support we needed
The needs assessment questions were not relevant to our jurisdiction's situation
Other, please specify:

```

```

Is there a standard procedure your team follows when developing your work plans?

Yes
No

Briefly describe the standard procedure your team follows when developing work plans:

```

```

Have you received support in developing your work plans from the technical assistance team?

     Yes
     No

If yes, did the support you received in developing your work plans align with the goals for your work?

     Yes
     No

If no, please explain why the assistance team wasn't helpful in supporting the development of your work plans.

Technical Assistance Technology Solutions

How has the TA impacted your communications between internal departments (HIV/STI)?
Select all that apply

     We work as a whole and share communication and data openly
     We are still working as separate departments and do not share data
     Data sharing has improved
     We have developed an SOP to help organize our process
     Other, please specify:

Have you been able to link HIV and STI data in a way that has allowed you to use this information in your work?

     Yes
     No

N/A

If no, what has been the challenge?

    The data sharing agreement has not been finalized
    There are challenges with the quality of data
    Other, please specify:

```
┌──────────────────────────────────────────────────────┐
│                                                      │
│                                                      │
│                                                      │
│                                                      │
│                                                      │
└──────────────────────────────────────────────────────┘
```

Why have you been unable to use linked HIV and STI data to conduct outreach to clients who have been out of care? *Select all that apply*

    Restrictions on in-person outreach
    Infrequent data matching
    Data sharing obstacles between surveillance and local health departments
    List for re-engagement is too long
    Other, please specify:

```
┌──────────────────────────────────────────────────────┐
│                                                      │
│                                                      │
│                                                      │
│                                                      │
│                                                      │
└──────────────────────────────────────────────────────┘
```

Approximately how many clients have you (your department) recorded as returned to care since using the linked data?

    < 10
    11-20
    21-50
    > 51
    N/A

Institutionalization of TA Interventions

Are there areas of technical assistance support that have become part of your daily routine?

     Yes
     No

If yes, briefly describe any areas of technical assistance support that have become part of your daily routine.

If no, why?

Do you feel you have the tools and resources needed to continue the work that has started as a result of this project, once the project period has ended?

     Yes
     No

If no, why?