REPORT OF THE VIRGINIA DEPARTMENT OF BEHAVIORAL HEALTH AND DEVELOPMENTAL SERVICES

Report on the Feasibility of Developing a Secure, De-identified, Renewable, and Relational Database of Criminal Justice, Behavioral Health, and Other Human Services Records (HJR 578, 2021 SSI)

TO THE GOVERNOR AND THE GENERAL ASSEMBLY OF VIRGINIA

HOUSE DOCUMENT NO. 6

COMMONWEALTH OF VIRGINIA
RICHMOND
2022

COMMONWEALTH of VIRGINIA

ALISON G. LAND, FACHE
COMMISSIONER

DEPARTMENT OF
BEHAVIORAL HEALTH AND DEVELOPMENTAL SERVICES
Post Office Box 1797
Richmond, Virginia 23218-1797

Telephone (804) 786-3921
Fax (804) 371-6638
www.dbhds.virginia.gov

January 17, 2022

Governor Ralph S. Northam
Patrick Henry Building
111 E Broad Street
Richmond, VA 23219

Dear Governor Northam:

HJ578 directs the Department of Behavioral Health and Developmental Services (DBHDS) to study the feasibility of developing a renewable, de-identified, relational database that links records from behavioral health, criminal justice, and other human services, for the purposes of developing more effective interventions.

The language states:

> In conducting its study, the Department of Behavioral Health and Developmental Services shall convene a work group composed of representatives of the Office of the Attorney General, Office of the Executive Secretary of the Supreme Court, the Departments of Health and Social Services, the Virginia State Police, the Virginia State Compensation Board, the Institute for Law, Psychiatry and Public Policy at the University of Virginia, and such other stakeholders as the Department of Behavioral Health and Developmental Services shall deem appropriate to determine the feasibility of establishing a secure, de-identified, renewable, and relational database that combines information from records created and maintained by criminal justice, behavioral health, and other human services agencies in the Commonwealth, which may be utilized by stakeholders to evaluate the effectiveness of existing interventions and facilitate the development of new interventions to improve outcomes. All agencies of the Commonwealth shall provide assistance to the Department of Behavioral Health and Developmental Services for this study, upon request.

In accordance with this item, please find enclosed the combined report for House Joint Resolution 578. Staff are available should you wish to discuss this request.

Sincerely,

Alison G. Land, FACHE

Commissioner
Department of Behavioral Health & Developmental Services

Sincerely,

Alison G. Land, FACHE
Commissioner
Department of Behavioral Health & Developmental Services

CC:
Vanessa Walker Harris, MD
Susan Massart
Mike Tweedy

1220 BANK STREET • P.O. BOX 1797 • RICHMOND, VIRGINIA 23218-1797
PHONE: (804) 786-3921 • FAX: (804) 371-6638 • WEB SITE: WWW.DBHDS.VIRGINIA.GOV

COMMONWEALTH of VIRGINIA

DEPARTMENT OF
BEHAVIORAL HEALTH AND DEVELOPMENTAL SERVICES

ALISON G. LAND, FACHE
COMMISSIONER

Post Office Box 1797
Fax (804) 371-6638
Richmond, VA 23218-1797

Telephone (804) 786-3921
www.dbhds.virginia.gov

Monday, January 17, 2022

The Honorable Janet D. Howell, Chair
Senate Finance Committee
14th Floor, Pocahontas Building
900 East Main Street
Richmond, VA 23219

Dear Senator Howell:

HJ578 directs the Department of Behavioral Health and Developmental Services (DBHDS) to study the feasibility of developing a renewable, de-identified, relational database that links records from behavioral health, criminal justice, and other human services, for the purposes of developing more effective interventions.

The language states:

> In conducting its study, the Department of Behavioral Health and Developmental Services shall convene a work group composed of representatives of the Office of the Attorney General, Office of the Executive Secretary of the Supreme Court, the Departments of Health and Social Services, the Virginia State Police, the Virginia State Compensation Board, the Institute for Law, Psychiatry and Public Policy at the University of Virginia, and such other stakeholders as the Department of Behavioral Health and Developmental Services shall deem appropriate to determine the feasibility of establishing a secure, de-identified, renewable, and relational database that combines information from records created and maintained by criminal justice, behavioral health, and other human services agencies in the Commonwealth, which may be utilized by stakeholders to evaluate the effectiveness of existing interventions and facilitate the development of new interventions to improve outcomes. All agencies of the Commonwealth shall provide assistance to the Department of Behavioral Health and Developmental Services for this study, upon request.

In accordance with this item, please find enclosed the combined report for House Joint Resolution 578. Staff are available should you wish to discuss this request.

Sincerely,

Alison G. Land, FACHE
Commissioner
Department of Behavioral Health & Developmental Services

Sincerely,

Alison G. Land, FACHE
Commissioner
Department of Behavioral Health & Developmental Services

CC:
Vanessa Walker Harris, MD
Susan Massart
Mike Tweedy

COMMONWEALTH of VIRGINIA

DEPARTMENT OF
BEHAVIORAL HEALTH AND DEVELOPMENTAL SERVICES

ALISON G. LAND, FACHE
COMMISSIONER

Post Office Box 1797
Fax (804) 371-6638
Richmond, VA 23218-1797

Telephone (804) 786-3921
www.dbhds.virginia.gov

Monday, January 17, 2022

The Honorable Luke E. Torian, Chair
House Appropriations Committee
13th Floor, Pocahontas Building
900 East Main Street
Richmond, VA 23219

Dear Delegate Torian:

HJ578 directs the Department of Behavioral Health and Developmental Services (DBHDS) to study the feasibility of developing a renewable, de-identified, relational database that links records from behavioral health, criminal justice, and other human services, for the purposes of developing more effective interventions.

The language states:

> In conducting its study, the Department of Behavioral Health and Developmental Services shall convene a work group composed of representatives of the Office of the Attorney General, Office of the Executive Secretary of the Supreme Court, the Departments of Health and Social Services, the Virginia State Police, the Virginia State Compensation Board, the Institute for Law, Psychiatry and Public Policy at the University of Virginia, and such other stakeholders as the Department of Behavioral Health and Developmental Services shall deem appropriate to determine the feasibility of establishing a secure, de-identified, renewable, and relational database that combines information from records created and maintained by criminal justice, behavioral health, and other human services agencies in the Commonwealth, which may be utilized by stakeholders to evaluate the effectiveness of existing interventions and facilitate the development of new interventions to improve outcomes. All agencies of the Commonwealth shall provide assistance to the Department of Behavioral Health and Developmental Services for this study, upon request.

In accordance with this item, please find enclosed the combined report for House Joint Resolution 578. Staff are available should you wish to discuss this request.

Sincerely,

Alison G. Land, FACHE
Commissioner
Department of Behavioral Health & Developmental Services

CC:
Vanessa Walker Harris, MD
Susan Massart
Mike Tweedy

1220 BANK STREET • P.O. BOX 1797 • RICHMOND, VIRGINIA 23218-1797
PHONE: (804) 786-3921 • FAX: (804) 371-6638 • WEB SITE: WWW.DBHDS.VIRGINIA.GOV

Virginia Department of Behavioral Health & Developmental Services

HJ 578

January 12, 2022

DBHDS Vision: A Life of Possibilities for All Virginians

1220 BANK STREET • P.O. BOX 1797 • RICHMOND, VIRGINIA 23218-1797
PHONE: (804) 786-3921 • FAX: (804) 371-6638 • WEB SITE: WWW.DBHDS.VIRGINIA.GOV

Preface

HJ578 directs the Department of Behavioral Health and Developmental Services (DBHDS) to study the feasibility of developing a renewable, de-identified, relational database that links records from behavioral health, criminal justice, and other human services, for the purposes of developing more effective interventions. The language states:

WHEREAS, the analysis of linked and de-identified information contained in criminal justice, behavioral health, and other human services records can facilitate a better understanding of service trajectories and long-term outcomes of various interventions; and

WHEREAS, understanding these outcomes can help guide policy making, resulting in more effective and efficient interventions for individuals involved in the criminal justice, behavioral health, and human services systems; and

WHEREAS, development of a coordinated state strategy concerning the collection and analysis of information in such linked and de-identified information contained in criminal justice, behavioral health, and other human services records may facilitate such collection and analysis, the development of interventions, and the formulation of policies; now, therefore, be it

RESOLVED by the House of Delegates, the Senate concurring, That the Department of Behavioral Health and Developmental Services be requested to study the feasibility of developing a secure, de-identified, renewable, and relational database of criminal justice, behavioral health, and other human services records to facilitate the development of more effective interventions.

In conducting its study, the Department of Behavioral Health and Developmental Services shall convene a work group composed of representatives of the Office of the Attorney General, Office of the Executive Secretary of the Supreme Court, the Departments of Health and Social Services, the Virginia State Police, the Virginia State Compensation Board, the Institute for Law, Psychiatry and Public Policy at the University of Virginia, and such other stakeholders as the Department of Behavioral Health and Developmental Services shall deem appropriate to determine the feasibility of establishing a secure, de-identified, renewable, and relational database that combines information from records created and maintained by criminal justice, behavioral health, and other human services agencies in the Commonwealth, which may be utilized by stakeholders to evaluate the effectiveness of existing interventions and facilitate the development of new interventions to improve outcomes. All agencies of the Commonwealth shall provide assistance to the Department of Behavioral Health and Developmental Services for this study, upon request.

The Department of Behavioral Health and Developmental Services shall complete its meetings by November 30, 2021, and shall submit to the Governor and the General Assembly an executive summary and a report of its findings and recommendations for publication as a House or Senate document. The executive summary and report shall be submitted as provided in the procedures of the Division of Legislative Automated Systems for the processing of legislative documents and reports no later than the first day of the 2022 Regular Session of the General Assembly and shall be posted on the General Assembly's website.

Table of Contents

Executive Summary

The current workgroup was established by HJ578 to study the feasibility of developing a de-identified, relational database that links records of individuals involved in the criminal justice system, who have also received behavioral health or other human services. The goal of such a system is to better understand the service needs and trajectories, along with the outcomes of various interventions, to inform future policy and program development.

The workgroup heard about efforts in Washington State to conduct similar work, and also learned about existing systems within the Commonwealth of Virginia that could support these efforts. One effort in particular, the Commonwealth Data Trust (CDT), was presented as an option for effectively linking data across different state agencies and organizations. The CDT is already fully supported by the Commonwealth, and has the ability to link data from multiple agencies to create a de-identified data set that can be utilized for further analysis. While the CDT was presented as a viable option for achieving the objective of HJ578, additional investments will be necessary as well. These include, financial investments in upgrading information systems that collect data, as well as sufficient staff to work with and maintain this data; further analysis of potential legal barriers toward sharing specific data sets; and formation of an oversight or steering committee, supported by agency and secretariat leadership, to guide these efforts.

Background

HJ578 directs the Department of Behavioral Health and Developmental Services (DBHDS) to study the feasibility of developing a renewable, de-identified, relational database that links records from behavioral health, criminal justice, and other human services, for the purposes of developing more effective interventions.

HJ 578 was enacted by the General Assembly in direct response to the increasing number of persons with serious mental illness (SMI) involved in the criminal justice system. This trend is a matter of deep and escalating concern. It is at the root of diverse and numerous policy challenges ranging from the proper role of law enforcement in mental health crisis response; increasing numbers of defendants whose competence to stand trial is questioned; the increasing waiting period for assessment, adjudication and treatment; the increasing proportion of state hospital beds being used for forensic patients; and the lack of community services to provide mental health services and supports to persons with serious mental illness in the justice system. The problems just described focus only on part of the story -- persons who are in the criminal justice system; a parallel set of challenges and frustrations arises for persons with SMI who have been diverted to the civil system for treatment but who require specialized interventions that are unavailable in the services system.

The growing intersection between mental health interventions and services and criminal justice

interventions and services highlights the fundamental point that mental health agencies and criminal justice agencies cannot respond to the changes they face, design interventions or policies, or assess their effectiveness without data maintained by both systems. That is an inescapable truth about mental health and criminal justice policy.

Efforts to evaluate service needs and outcomes at the intersection of behavioral health, social services, and the criminal justice system require coordination across many different local and state agencies. At the state level alone, behavioral health, social services, and health data may be contained at DBHDS, the Department of Social Services (VDSS), Adult Protective Services (APS), the Department of Medical Assistance Services (DMAS), or the Department of Health (VDH). Similarly, data on individual encounters with the criminal justice system may occur with the Virginia State Police (VSP), the Department of Corrections (DOC), Department of Criminal Justice Services (DCJS), the Department of Juvenile Justice (DJJ), the Executive Secretary of the Supreme Court (OES), the Virginia State Compensation Board (VSCB), as well as local police, regional jails, and others. Any attempt to link data across these systems requires specific data sharing agreements from each agency to every other agency that will have access to their data. Consequently, attempts to link data from three or more agencies becomes exponentially more complex as more agencies are included. The effort involved in creating such agreements is a significant barrier to performing this analysis.

Once data sharing agreements have been established, the technical work of cleaning and curating each data set, the various data sources must be linked on the basis of some common identifier(s), and uncertain matches must be reviewed and resolved before the complete data set can be analyzed. Linking these multiple data sets must be handled by a single agency or a separate organization contracted to do this work. Historically performing each of these necessary tasks has been complicated and time consuming and served as a significant barrier to performing these analyses and furthering our understanding of how to improve services within this area. The current workgroup was established to determine whether it is feasible to develop a more sustainable method of linking data from these various systems, in order to better understand the needs and outcomes of individuals involved in the criminal justice and behavioral health systems.

Washington State

The workgroup heard a presentation from David Mancuso, Ph.D., the director of the Washington State Department of Social and Health Services, Research and Data Analysis (RDA) Division. Dr. Mancuso reported on the work that has been done in Washington State in developing the Integrated Client Databases (ICDB), which includes services data across several state agencies that include data from behavioral health, child welfare, courts, the state patrol, medical assistance, and the departments of health, corrections, and commerce, among others.

Dr. Mancuso noted that a significant number of individuals receive services across multiple agencies, and this is a key rationale for integrating data across these systems. When looking at individuals using services across three agencies in Washington, the RDA found that 60-80 percent of the individuals receiving services from one agency were also receiving services from one or more additional agencies. This means that using data, siloed within a single agency limits

our understanding of how individuals utilize services, what risk factors in one area (e.g., mental health) increase the likelihood of negative outcomes in another area (e.g., criminal justice), and how supports or services in one area may lead to better outcomes in another.

Dr. Mancuso presented several examples of analyses conducted by the RDA that utilized linked data from multiple state agencies to facilitate policy analysis, evaluate programs, and measure performance and outcomes. Many of the examples presented utilized Medicaid claims linked with other agency data to understand how risk factors, such as chronic physical conditions, poverty, mental illness are related to the utilization of services and health outcomes. Other work has identified the support needs common to high cost beneficiaries and the types of interventions that can reduce the risk of criminal justice involvement.

Framework for Addiction Analysis and Community Transformation (FAACT)

Within the Commonwealth, the Framework for Addiction Analysis and Community Transformation (FAACT) is a recent example of using data linkages across agencies to help local communities respond more effectively to the opioid public health crisis. FAACT involved collaboration between over 20 state and local agencies involved in addressing various aspects of the opioid epidemic, including: law enforcement, emergency services, behavioral health, private hospitals, community jails, and the court system.

Rather than a statewide program, FAACT was developed and driven from the community level. The initial project was developed through a partnership of state and local agencies providing services in the Northern Shenandoah Valley, and later expanded to the Roanoke Valley and Southwest Virginia. With the assistance of Qlarion, the Northern Shenandoah Valley Substance Abuse Coalition, along with the Virginia Department of Criminal Justice Services made data from the various agencies available to the Virginia Commonwealth Data Access Portal, which was then used to obtain a broader picture of the impact of the opioid crisis and connections with other services.

The coalition used data obtained through the FAACT platform to inform actions to address the opioid epidemic in their community. This included:

- Using data on where overdoses and drug related criminal activity was occurring to target deployment of staff to where they were most needed
- Utilizing data on first use to increase youth engagement in evidence based services
- Identifying emerging drug threats early

FAACT serves as a model for how the goals ("…developing a secure, de-identified, renewable, and relational database of criminal justice, behavioral health, and other human services records to facilitate the development of more effective interventions.") of HJ578 can be achieved. While the data utilized for FAACT was largely regional, rather than statewide, it serves as a model for how such a statewide data system could be established. Data contributed to FAACT included information from criminal justice (pre-trial, community corrections, police incidents), health (regional overdoses, lab results, hospital admissions), and other human service records (social services, unemployment, poverty, education). While it did not include behavioral health data; that could easily be added.

Uses of Linked Data

Aggregate data from separate data systems can be used to understand trends in behavioral health and criminal justice issues, looking at patterns over time or across geographic regions. For example, data showing increases in drug use coincides with increases in criminal arrests can be used to infer a relationship between the two. However, it is difficult to know whether indeed the data are linked, or perhaps both responding to some other, unknown, factor. Integrating the data between various systems can provide a better picture of the needs of specific groups of individuals, as well as the effectiveness of interventions designed to address these needs. Some examples of the types of analyses that could result from a linked data system include:

- Evaluating interventions designed to reduce incarceration of people with mental illness are most effective?
- What is the outcome for those who are found not guilty by reason of insanity (NGRI); do they receive appropriate services; what is rate of being incarcerated in the future?
- What is the role of policing practices on response to mental health crises? What is the relation between emergency mental health intervention and occurrence of an arrest?
    - How many charges of felony assault on a law enforcement officer were related to mental health crises?
    - Are there other charges that appear frequently during the period before or after an individual receives an ECO/TDO?
- Improve our understanding of how individuals move through the behavioral health and criminal justice system; evaluating services at different intercepts with the criminal justice system[1]:
    - How many people are diverted from jails to community based crisis services; what services are requested?
    - How many people are diverted from jail (or some other higher level of care) by mobile crisis services; what interventions are most effective?
    - What proportion of calls to 911 are related to behavioral health issues; what is impact of the Marcus Alert legislation and the new crisis service model?
    - How many people have competency issues raised before trial; how long does it take for evaluations to occur; how long do people found not competent remain in jail before being transferred for treatment; what is the impact on jails?
    - How do social determinants of health impact behavioral health and criminal justice involvement?
- What are the costs of treating people with behavioral health conditions in the criminal justice system?
- What types of disparities are present in services across agencies by race, ethnicity, disability, geographic location, etc.?
- Do individuals involved with the criminal justice system have different outcomes when they also receive services from behavioral health, or other service agencies?
- Do individuals who are ordered to mandatory outpatient treatment experience fewer crises, arrests or detentions incarcerations in the following 12 months? If so, is the reduction sustained at 24 months?
- Are their gaps in behavioral health or other services for certain populations, or in certain localities? What impact does this have on involvement in the criminal justice system?

[1] Substance Abuse and Mental Health Services Administration. *Data collection across the sequential intercept model.* August 2019. Available at https://store.samhsa.gov/sites/default/files/d7/priv/sma19-5097.pdf

Options for Linking Data

The Commonwealth Data Trust

The Commonwealth Data Trust (CDT), managed by the Office of Data Governance and Analytics (ODGA), under the Commonwealth Chief Data Officer, provides a means of sharing data and information in a safe, secure manner. The CDT was developed to provide a trust framework for the Commonwealth, enabling a common set of rules and polices for the purpose of information sharing by trust members. Rather than each agency establishing specific data use agreements with each agency they with whom they wish to share data, the CDT establishes standard data use and sharing agreements that govern the use of data. Each agency that chooses to participate in the CDT determines which specific data sets they would like to make available to the CDT and determines who can have access to that data. Data owners retain control of their data and make determinations on who their data can be shared with and for what purposes. Data owners may choose to share some, all, or none of their data. Agencies or organizations that participate in the CDT sign specific trust agreements, which include:

- Data Trust Member Agreement – establishes the relationship between the organization providing the data and the trustee of the data (ODGA).
- Data Trust User Agreement – describes the relationship between the trust and recipients of trust member-contributed restricted-use data. This agreement is primarily for organizations that have projects that need access to restricted-use trust data.
- CDO Onboard Readiness Review – used as a guideline prior to onboarding data to the CDT.

The CDT identifies five tiers of data to identify the level of privacy and confidentiality of the data. Four tiers are appropriate for inclusion in the CDT and range from tier 0, which indicates the data is neither sensitive nor proprietary and is intended for public use; to tier 3 data, which includes sensitive or proprietary information and data elements with a statutory requirement under Data Trust Member's relevant state and federal laws for notification to affected parties in case of a confidentiality breach. Examples of tier 3 data include social security number, driver's license number, or personal medical information. Tier 4 data, which should not knowingly be included in the CDT, includes data that is sensitive or proprietary data where the unauthorized disclosure could potentially cause major damage or injury, including death, to entities or individuals identified in the information, or otherwise significantly impair the ability of the Data Trust Member to perform its statutory functions.

Data that is shared through the CDT can be linked with other data sets, without exposing protected individually identifiable information (PII) or protected health information (PHI) in the linked data set. This process is managed through Data Sage, Virginia's Secure Analytics and Governance Environment, which uses an anonymized crosswalk (Universal Entity Index) from different data systems to create a master person identifier for each individual. This anonymized crosswalk table is used to match de-identified attribute data for an individual across multiple systems. Essentially personally identifiable attributes (name, date of birth, race, gender, etc.) is kept secure, while a universal identifier is used to link records across different systems. This

results in a de-identified table that can show the services used by an individual in one agency (e.g., vaccine administrations), with those utilized through another agency (e.g., use of SNAP benefits). This de-identified data can be used to better understand patterns of use, which can help understand the effectiveness of programs, as well as drive future interventions.

Virginia Longitudinal Data System (VLDS)

The Virginia Longitudinal Data System (VLDS) was originally funded by a grant from the United States Department of Education in 2009 and serves to link data across multiple state agencies to answer policy and research questions. The VLDS uses a federated model to link various data systems. This means that the private, confidential data remains behind each agency's firewall, while a complex hashing process is used to create a de-identified, linked data set between two or more agency records. Agencies participating in the VLDS include the Virginia Department of Education (VDOE), the State Council of Higher Education for Virginia (SCHEV), the Virginia Employment Commission (VEC), the Virginia Department of Social Services (VDSS), the Virginia Community College System (VCCS), the Virginia Department for Aging and Rehabilitative Services (DARS), and Virginia Department of Health Professions (DHP).

According to the VLDS website, "VLDS was designed to balance privacy against the ability to discover insights about Virginia's education, workforce, higher education, and health policies and programs. These insights can lead to increased agency performance, improved government policy, and improved citizen outcomes."

The VLDS is not a database or a data warehouse, but rather a system for linking data from participating agencies to address specific research or policy questions. The VLDS is organized around a set of burning questions and the VLDS research agenda, which is informed by its members and stakeholders. Researchers who wish to access data from the VLDS must identify an agency sponsor and work with them to develop research questions that are consistent with the burning questions and research agenda. The de-identified dataset produced by VLDS can never be traced back to a specific individual.

Diveplane

While not a solution for linking data, the committee also heard about Diveplan, a product being utilized by DBHDS to further secure PHI. Diveplane is an artificial intelligence (AI) product that is used to create a synthetic version of the data, which eliminates any PHI, while retaining the overall values and relationships between variables in the aggregate data set. Diveplan would create synthetic values for all variables in the data set (name, date of birth, address, etc.) such that the data could be shared without exposing confidential information. However, the overall data values are preserved, such that the aggregate synthetic data set should return the same overall results as the original data set. Thus, the full synthetic data set (as opposed to aggregate results) can be shared or released for further analysis without exposing PHI or PII. This creates the potential for further analysis by researcher or evaluators of certain data (such as geographic location) that may need to be redacted as PII, but may also have analytic value.

Barriers

Data Availability and Data Quality

Some data that may be of interest is captured at a local level and not available through a single state agency. For example, within the court system, clerks of court are the custodians of court records pursuant to Va. Code §§ 17.1-242 and 16.1-69.54. Circuit court clerks are elected constitutional officers with the authority to select and maintain their own case management systems and not all clerks use the Office of the Executive Secretary's circuit case management system. Therefore, making this data available would require working with multiple entities and multiple systems, which may not directly align with each other.

The quality of the data that would be important to agency staff, policy makers, and/or researchers varies across data sets and agencies. While some data sets are well defined and collected through interfaces with validation rules to minimize inaccurate data entry, others are not well documented and/or rely on outdated technology. For example, recent efforts by researchers at Duke and the University of Virginia to analyze data across six state agencies (including VSP, DMAS, DBHDS, VDH, DOC) encountered multiple issues with data quality, including unclear documentation about data fields and their meaning; as well as varying quality of identifiers, including misspelling of names, typographical errors in birth dates, etc. For example, information such as diagnosis and involuntary admission status was missing from many hospitalization records. Errors in identifier fields required a high level of manual review to attempt to address these errors and match records and resulted in some obvious mismatches; for example, event records that occurred after a person's death record date.

While the ODGA will work with agencies to document and clean up data prior to including it within the CDT, there are limits to how effectively this can improve the quality of flawed data. Ultimately, the ability to effectively link data across agencies will rely on each agency establishing data governance and data quality processes to ensure reliable and valid data in their primary source systems. Agencies need staff with expertise in data management and information technology; systems for collecting data should have methods for ensuring the validity of data when it is entered.

Unique Identifiers

Linking data across different data sets and across agencies requires common unique identifiers to allow the matching of a unique individual in one system with that same individual in another system. In theory, a single unique identifier, such as social security number, could be used as a common identifier to link data across agencies. This is known as deterministic matching. However, in practice this may not be effective, as some individuals may decline to provide a social security number, others may provide multiple different numbers when presenting for different services, in addition this data may be entered incorrectly due to human error.

Deterministic matching may work best in settings (such as a healthcare system) where there is an organization that is responsible for assigning a unique identification number for all individuals in that system. Currently there is not a single identifier that is used by all agencies within the Commonwealth, nor is there a single entity responsible for assigning identification numbers to

all individuals in the Commonwealth. This would likely require legislation and significant resources to implement.

In the absence of a single unique identifier, or in situations where information in identifiers may be incorrect, best practices for matching data relies on probabilistic matching. This involves comparing several different identifiers across different data sets to identify when there is a high probability that the records are from the same individual, even when they do not match exactly. For example, slight variations in the spelling of a name (Michele vs Michelle) might be considered the same person if other identifiers (gender, date of birth, address, etc.) do match. When conducting probabilistic matching, the more identifiers that are shared in common between two data files, the better the ability to correctly identify matching records.

Data Use Agreements and Legal Issues

Currently, any effort to link data from different agencies requires a specific data sharing agreement that specifies the specific data that will be shared, how it will be used, and the protections that each party will take to ensure the confidentiality of the data. Should those agencies desire to link additional data, or to link with data from a third agency, additional data sharing agreements would need to be executed. This process is typically time consuming and is often a barrier to quickly sharing data in real time when as emerging issues arise (such as the COVID-19 pandemic). In the above-mentioned Duke/UVA study, obstacles with executing data sharing agreements between all participating agencies took much longer than anticipated and resulted in the grant funding expiring before the study could be completed.

The CDT serves to eliminate the need for individual data sharing agreements between agencies, by creating an overall trust framework across all members. All agencies agree to specific requirements for sharing and use of data. Each agency decides which specific data that would like to make available to the CDT, whether all fields or a limited set of fields will be included, as well as the sensitivity level of their data. Each agency also retains control of data that they have made available to the CDT and can continue to determine who can have access and for uses of the data. As such, the CDT streamlines the sharing of data between agencies.

Because data that is linked has individual identifiers removed prior to the linked data set being made available, many legal issues around confidentiality should be addressed. However, there are many different federal and state laws regarding the allowable use of data, therefore each agency would need to make determinations with their legal counsel as to their ability to make certain data available to others. Agencies participating in the workgroup had differing understanding of their ability to legally include information in the CDT. Further review by each agency's legal counsel, or collectively by the Office of the Attorney General would be needed to further clarify these issues. Continued efforts working across Secretariats and Agency Leaders will further drive adoption and value extraction and consistency across the Commonwealth.

Privacy/Confidentiality

HJ578 specifically addresses studying the feasibility of creating a "de-identified" database. However, linking data across different data systems and different agencies requires that

individual identifiers are used to create the initial link between data. The workgroup stressed the importance of ensuring that any data that is produced and shared has all personal identifiers removed and cannot be traced back to a specific individual.

Resources

Addressing many of the barriers identified above will require the investment of significant resources by the Commonwealth. For example, the Office of Forensic Services, within DBHDS utilizes an application called the Forensic Information Management System (FIMS) to track forensic hospital admissions and Not Guilty by Reason of Insanity (NGRI) patients who receive services in the hospital and in the community. This program was initially developed in 1992 as a Microsoft Access database and needs significant modernization. Reporting from the FIMS system (upcoming court dates, evaluation dates, expiration of court orders, etc.) is accomplished through a companion application called FIMS Reports, which are tracked manually through a Microsoft Excel spreadsheet. This redundancy and manual entry of data all increase the likelihood that user error may result in inaccurate data. Ideally, these systems need to be updated to modern technical and security standards prior to inclusion in a state-wide data ecosystem. It will be important that each agency have subject matter experts that are knowledgeable about the specific data maintained by that agency, including the overall quality of the data, the potential limitations of that data, and an understanding of how to interpret the data.

In addition to the technical resources required to modernize some of the applications that collect and store relevant data, workgroup members identified the need for additional staff resources. Additional staff would be important to facilitate collecting, maintaining, and curating data within agencies; as well as to engage in cross-agency discussion identifying policy or evaluation questions, the relevant data sources that can be used to address these questions, and reviewing results of such studies to ensure that conclusions are consistent with the data itself.

Focus

There was general consensus across members of the workgroup that some level of sharing and integrating data would be beneficial and could lead to new insights and understanding about individuals interactions and outcomes across the behavioral health and criminal justice system. However, there were not any unifying or emerging themes around which the group was able to coalesce. Given the resources that will be necessary to develop a de-identified data system that links data from the behavioral health and criminal justice systems, having a common purpose, with questions important to each agency will be important to sustain the level of effort that will be needed to develop and maintain such a data system.

Recommendations

Based on workgroup meetings and additional research, the following recommendations are made:

1. In order to accomplish the intent of HJ578, it is essential that each agency has adequate resources for the task. While evaluating individual data sets and resources at each agency

was beyond the scope of this workgroup, it is clear that additional agency resources will be needed to establish a meaningful system of data that can provide valid and reliable information on program outcomes. Some of the data sources are outdated and require procuring new systems to collect accurate information that can be reliably linked with other data sources. In addition, it is necessary that agencies have staff that are knowledgeable about their data sources and are able to work with other agencies and the CDT to ensure that any findings from analyses are interpreted appropriately. A commitment to necessary funding is a necessary condition for moving forward.

2. Because this effort would be cross-Secretariat, Executive Branch leadership should ensure the necessary buy-in for the described data linkage and analysis from the agencies most concerned with addressing the needs of persons who regularly encounter the behavioral health, public safety, and criminal justice systems and who move between these systems over time. Agencies should identify individuals within their agencies to coordinate efforts with other agencies to secure the interagency agreements and share the necessary data. Leadership should identify and deploy the resources needed to build individual data sets of sufficient quality to allow linking with other data sets for policy evaluation. Leadership should identify within-agency information technology experts to guide the process and to facilitate secure data preparation and transfer. Leadership should also identify appropriate regulatory personnel to ensure compliance with applicable laws for ensuring privacy of individual records.

3. The Commonwealth should establish an ongoing structure that supports the linking, assembling, and analyzing administrative record data across agencies. This could take the form of a cross-agency workgroup to formulate and prioritize burning questions, identify data sources that will address those questions, identify and address technical challenges, and ensure the correct use and interpretation of any cross-agency data linkages that occur. Identification of data sources will need to consider data that is available state-wide through a single agency, and data that is collected by local agencies and not available through a single system. Such a steering committee or workgroup should be supported by agency and secretariat leadership to ensure allocation of appropriate resources. The workgroup should also include stakeholders with appropriate knowledge and expertise to guide the work, including researchers and other social scientists, community advocates, and others.

4. Consider leveraging the work that has already been started through FAACT. This collaboration, driven by local and state-wide organizations, has successfully used data from the criminal justice, health, and social services areas to generate actionable information. Although much of the data utilized for FAACT was generated at the local level, this could serve as a starting point for a more comprehensive, state-wide system.

5. Continue to explore potential legal issues and limitations in sharing agency data through the CDT. Several agencies identified legal restrictions on sharing their data as a barrier to participating in the CDT. However, because agencies participating in the CDT do not directly share data and maintain control over who and what others are able to access, the ODGA has stated that much of these concerns can be addressed through the Trust

agreements. According to the ODGA, the CDT utilizes a Universal Entity Index crosswalk table that can link individuals across multiple systems without using PII. The technical details of this process, along with the Trust agreements will need to be explored further by each agency's IT staff and legal counsel to identify and address specific concerns.

Conclusion

The Commonwealth is currently engaged in efforts that will readily support efforts to link records and information from the behavioral health, criminal justice, and other health and human service records. The VLDS is a system that has been developing over the past ten years and includes records from VDOE, SCHEV, VEC, VDSS, VCCS, DARS, and DHP. This system is primarily focused on education, workforce, and health policies; though it could potentially be leveraged to include behavioral health and criminal justice information. The VLDS is organized around a set of "burning questions" and the research agenda of the member agencies; which guide the analyses that are conducted on the data.

The other effort, operating under the ODGA is the CDT, which was established to provide a secure environment for sharing data and information across agencies and organizations throughout the Commonwealth. The CDT serves as the trustee for data and information that is shared by member organizations, and is responsible for ensuring the security and privacy of that information. Member organizations retain control over who is able to have access to their data, however, the trust framework provides a structured and streamlined format for user sharing agreements. Data from that is made available to the CDT by different organizations can be linked together and the combined data set can be shared (without identifying information) for further analysis to address policy questions, or evaluate program outcomes.

While neither the CDT or the VLDS are technically a database, they are both platforms that facilitate the sharing and linking of data; they would not require defining all of the tables and relationships in advance, as one might do with a database; and would have greater flexibility to add additional data and scale up or down as needs or interests dictated. In addition, they both work to preserve privacy and confidentiality, since neither requires sharing PII or PHI with other agencies.

Although both the VLDS and CDT could serve as a mechanism for sharing and linking data, the VLDS has been focused around education and employment data, and therefore may not be a natural fit for including criminal justice and behavioral health information. The CDT has not been organized around any particular themes or content, and is open to data from any organization in the Commonwealth that agrees to become a member of the Trust. The CDT has already been used to support data sharing and analyses for FAACT, and this experience could be built upon to develop a more extensive framework for sharing and linking criminal justice and behavioral health data. The CDT can provide an effective framework for linking behavioral health and criminal justice data, however, on its own, it lacks a set of organizing principles that would help guide the developing and inclusion of data and the resulting analyses. The VLDS is organized around a set of "burning questions" developed by the member organizations, as well

as by the research agendas of each of the individual agencies. A similar framework would be helpful to guide the development of system for linking behavioral health, criminal justice, and other social service records.

However, it must be noted that carrying out the intent of HJ578 would need to overcome significant barriers including data availability and data quality, data use agreements and legal issues, privacy/confidentiality, funding and resources and achieving a single focus for the many agencies involved. If the will is there, the recommendations in this report can help to overcome barriers to developing a renewable, de-identified, relational database that links records from behavioral health, criminal justice, and other human services, and so develop more effective interventions for people served by the Commonwealth.

Acknowledgements

DBHDS would like to extend its appreciation to the members of the workgroup and to all of those who provided assistance for this study. We would also like to offer a special thank-you to those who presented at the workgroup:

David Mancuso, Ph.D.
Director, Washington State Department of Social and Health Services
Research and Data Analysis Division

Marcus Thornton
Deputy Chief Data Officer / Interim Chief Data Officer
Office of Data Governance and Analytics
Commonwealth of Virginia

Alan Cross
Chief Commercial Officer
Diveplane Corporation

Appendices

Appendix A – Workgroup Participants and Meeting Dates

Aaron Bloomfield (University of Virginia)
Richard Bonnie (University of Virginia, ILPPP)
Bruce Cruser (Mental Health America of Virginia)
Catherine Day (Virginia State Police)
Heidi Dix (Department of Behavioral Health and Developmental Services)
Michele Easter (Duke University)
Diane Enroughty (Department of Medical Assistance Services)
Jennifer Faison (Virginia Association of Community Services Boards)
Rashmi Gupta (Department of Medical Assistance Services)
Ashley Harrell (Department of Medical Assistance Services)
Alex Der Hovhannesian (Clarion)
Kari Jackson (State Compensation Board)
Mike Jagels (Office of the Attorney General)
Mike Jones (Department of Medical Assistance Services)
Brian Lacks (Virginia State Police)
Jason Lowe (Department of Medical Assistance Services)
Anna Mendez (Partner for Mental Health)
Colleen Miller (Disability Law Center of Virginia - dLCV)
Sraven Morsheti (Department of Behavioral Health and Developmental Services)
Dev Nair (Department of Behavioral Health and Developmental Services)
Bonnie Neighbour (VOCAL)
Toyin Ola (Department of Behavioral Health and Developmental Services)
Heather Orrock (VOCAL)
Alisa Padden (Supreme Court of Virginia/OES)
Mathew Paterson (Virginia State Police)
Jeff Price (Department of Social Services)
Mike Riggs (Supreme Court of Virginia/OES)
Chad Rogers (Virginia State Police)
Christine Schein (Department of Behavioral Health and Developmental Services)
Glenn Schmitz (Department of Behavioral Health and Developmental Services)
Jess Schneider (Department of Juvenile Justice)
Jeff Swanson (Duke University)
Angela Torres (Department of Behavioral Health and Developmental Services)
Allyson Tysinger (Office of the Attorney General)
Paul Whitchurch (Department of Behavioral Health and Developmental Services)
Craig White (Virginia State Police)
Christie Wu (Department of Medical Assistance Services)
Heather Zelle (UVA)

Meeting Dates

September 21, 2021
1) Welcome and Introductions
2) Charge for the Workgroup

3) Value of Integrated Data – Washington State
   David Mancuso, PhD, Director, Research Data Analysis
   Washington State Dept. of Social and Health Services
4) Commonwealth Data Trust
   Marcus Thornton, Deputy Chief Data Officer
   Office of Data Governance and Analytics

November 9, 2021

1) Introductions
2) Agency Discussion
3) Uses for linked data
4) Potential data
5) Resources Needed
6) Barriers

November 30, 2021

1) Welcome/Introductions
2) Review of Prior Meetings
3) Diveplane
4) Additional Questions/Concerns