Best Practices in Data Integration

Compiled for the Vermont Healthcare Workforce Data Center (VT HWDC)

April 10, 2024

Prepared by Freedman HealthCare

Table of Contents

EXECUTIVE SUMMARY

Vermont's HealthCare Workforce Data Center (HWDC): Supporting Healthcare Reform Workforce Development

In 2022, the Vermont General Assembly took a crucial step forward in addressing the growing shortage in the healthcare workforce by passing Section 32 of Act 183 which appropriated funds to establish and maintain an integrated Health Workforce Data Center (HWDC). The intent of this legislation and the development of the HWDC is to:

- Combine available national data as well as various data sources related to Vermont's healthcare workforce to support the State's ability to assess its current and future workforce needs.
- Identify opportunities for workforce education, training, and development.
- Provide modeling and analytic reporting that will allow health policy makers, health care employers, health care employees, and health care students to make meaningful and informed decisions about their role in supporting and developing the health care workforce.

In late 2023, the Vermont Agency of Human Services contracted with Freedman Healthcare, LLC (FHC) to work collaboratively with the HWDC manager and provide consulting services in the development and planning of the HWDC which encompasses three phases:

- Phase 1: Discovery and planning
- Phase 2: Stakeholder engagement
- Phase 3: Analytics and operations framework

As part of the first phase, Discovery and Planning, FHC provided 8 sessions (see *Appendix A*) on integrated data systems (IDS) best practices, reviewing an overview of other IDS initiatives, data governance, data architecture, and funding and sustainability. Two of the sessions included representatives from the fully implemented RI Data Ecosystem and the in-development IDS in Hawaii, Hawaii Analytics Program (HAP), who provided insights, considerations, and lessons learned from their first-hand experiences.

Overview of integrated data systems.

The state of Vermont can benefit from the experience of other states that have implemented integrated data systems and/or have collected healthcare workforce data. An overview of similar programs in Rhode Island, Delaware, Hawaii, Washington, California, and Virginia covered the areas of data governance, data sources contributed, program funding, and resulting data products and reports.
Key takeaways include:

- Integrated Data Systems need time to mature.
- Start small, use a phased approach to allow time for testing, flexibility to adjust as needed, and working through new processes, while continuously building towards the IDS end goal.

- Communicate effectively and efficiently with team members, stakeholders, partners, and other interested parties to develop a productive and collaborative environment.
- Develop and maintain positive relationships and data sharing agreements between state agencies.

Data Governance. Data Governance is a set of policies, procedures, and related documentation that guides how people will manage data to create actionable information and maximize value. Broadly, data governance manages who will access what data, when, and under what circumstances. Applied to an Integrated Data System (IDS), data governance is the practice of knowing where the data came from, how it will be utilized, and how it will be adequately protected. A successful data governance program:

- Identifies and engages teams, leaders, and stakeholders that will support both the vision and the operations of the IDS.
- Defines the data governance roles and responsibilities.
- Forms an effective data governance team by providing orientation, education, and training.
- Builds trust and a collaborative environment by providing effective and frequent communications between the data governance team and its stakeholders.
- Establishes procedures to safeguard the data throughout the process, from data collection to data release.

Data Architecture. Data architecture focuses on the way an organization's policies, standards, and rules are structured to fulfill an organization's business needs. In the case of IDSs, business needs often call for the integration of datasets. A successful integrated data architecture will include an assessment of existing infrastructure, the implementation of meaningful metadata, and the integration of data governance systems to consistently monitor the lifecycle and quality of data assets. A summary of data architecture key takeaways and considerations include:

- Business needs should drive technological decisions.
- Data security and protection are critical.
- Effective systems include leveraging enterprise resources for managing master data and metadata. Being able to integrate with resources such as a master person index or provider directory will assist with keeping integrated data current and improve its quality.
- Evaluating opportunities to extend or reuse existing architecture should be done carefully, considering regulatory compliance and modularity.
- An integrated data system is built on trust and the architecture should embed quality assurance and security throughout the data lifecycle.

Funding and Sustainability. Integrated Data Systems should consider all available and prospective revenue sources to support their initial development and implementation, and then ongoing operations thereafter. While the VT HWDC has relied solely on a state appropriation to embark on the planning and design phase of the work, this funding source may not be sufficient to cover all program costs in perpetuity. Developing a sustainable and diversified funding plan at the outset of the HWDC will provide stability to the project team and promote stakeholder buy-in and engagement in a longstanding program. Potential data sources include state appropriations, revenue-generating data release products, assessments or ongoing subscriptions for data access, federal funds, and private grants, trusts, and foundations. It is

recommended that the State explore funding options early as the processes involved in accessing the funds may take longer than expected.

INTRODUCTION

Vermont's Healthcare Workforce Data Center (HWDC)

In 2022, the Vermont General Assembly passed Section 32 of Act 183, which appropriated funding for the development of a statewide Healthcare Workforce Data Center (HWDC). This integrated HWDC will combine available national data as well as various data sources related to Vermont's healthcare workforce to support the State's ability to assess its current and future workforce needs; Identify opportunities for workforce education, training, and development; and provide modeling and analytic reporting that will allow health policy makers, health care employers, health care employees, and health care students to make meaningful and informed decisions about their role in supporting and developing the health care workforce.

To advance the goal of designing and planning for HWDC, the Vermont Agency of Human Services contracted with Freedman HealthCare, LLC (FHC) in late 2023 to provide subject matter expertise, summarize best practices, and convene stakeholders to explore how to most responsibly, efficiently, and economically establish the HWDC and develop a meaningful analytics program. The final planning deliverable is a Five-Year Operations Plan that includes staffing needs as well as a detailed project plan to operationalize the development of a robust HWDC that meets the requirements of Act 183 and is efficient, responsive, and scalable. In addition, the FHC will develop a Reporting and Analytics Plan that will outline a preliminary approach for developing "first tier" reports using HWDC data in the short- and medium-term.

To build a shared understanding of the features, structures and development processes for HWDC, FHC facilitated eight "Best Practices Meetings" with Vermont administrators to explore integrated data system (IDS) initiatives in Rhode Island, Hawaii, and Delaware and key elements in data governance, data architecture, staffing considerations, and funding models.

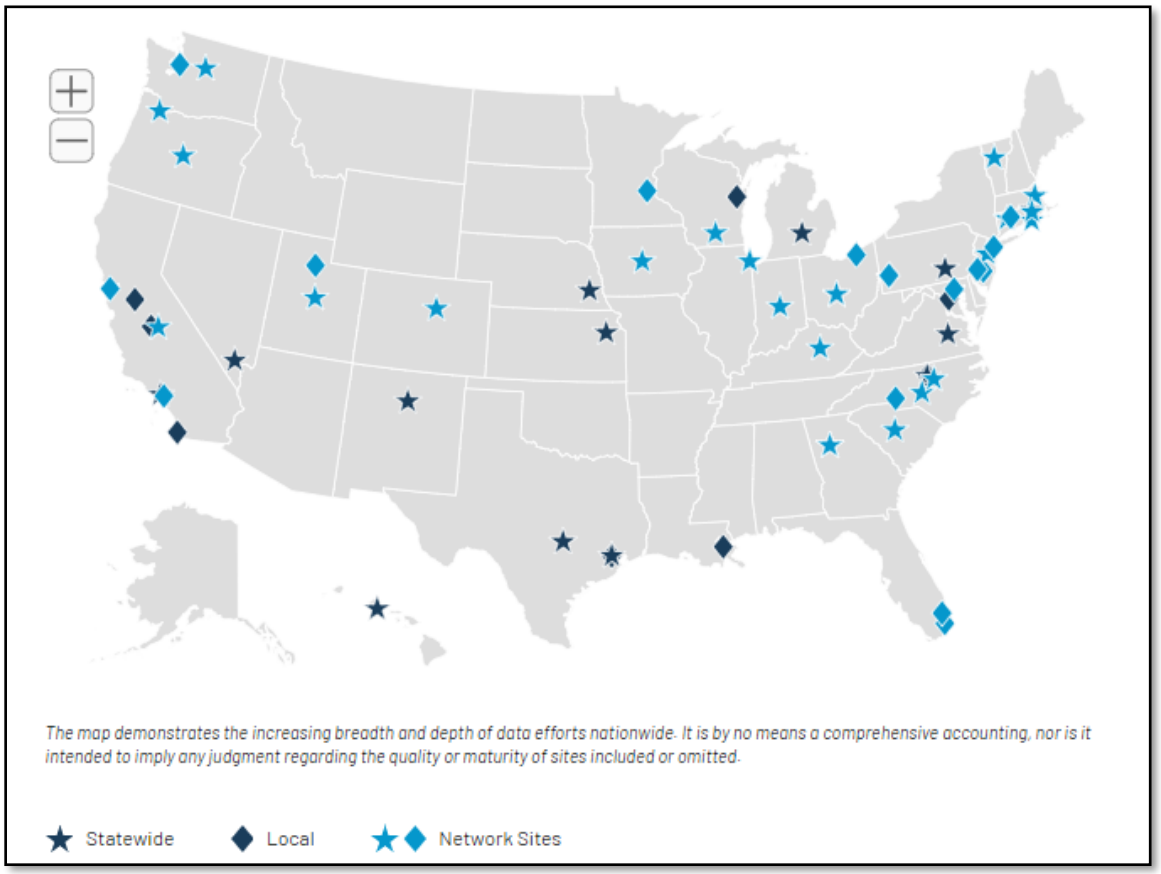
This report summarizes the information shared during these meetings and serves to provide background and reference information as the project moves forward. Readers may also reference the slides from the Best Practices meetings. as well as the list of speakers and topics in Appendix A.

CHAPTER1: OVERVIEW OF INTEGRATED DATA SYSTEMS

Introduction

Integrated data systems (IDS) are centralized repositories of administrative data that combine information across various sources to provide a unified, more complete portrait of an individual or cohort and their various touch points across multiple agencies' services, programs, and/or dimensions. Data Hubs, Ecosystems, Data Collaboratives, and Data Intermediaries are other names for IDSs. Regardless of what they are called, successful efforts to leverage shared data to improve individual and community outcomes must address similar data governance, ethical, relational, legal, and technical considerations. Figure 1 below provides a snapshot of the data sharing and integration efforts across the nation.

Figure 1 - Actionable Intelligence for Social Policy (AISP) Integrated Data Systems[1]



[1] Actionable Intelligence for Social Policy (AISP), UPenn, 2024 https://aisp.upenn.edu/integrated-data-systems-map/

Key Takeaways:

- Take advantage of existing, similar human services IDS in evaluating potential models for VT.
- Data systems need time to mature.
- Start small, with a phased approach to allow time for testing, maintaining flexibility for necessary adjustment, and working through new processes while continuously building towards the IDS end goal.
- Communicate effectively and efficiently with team members, stakeholders, partners, and other interested parties to develop a productive and collaborative environment.
- Develop and maintain positive relationships and data sharing agreements between state agencies.

Data Sharing vs Data Integration

Data sharing refers to providing other parties with access to data not otherwise available to them. Data integration is a more complex type of data sharing that involves joining or merging different data sources using common fields to provide users with a unified view of the data. Common fields used to link data sources at a person level include direct identifiers such as Social Security numbers (SSN), names, dates of birth, National Provider IDs (NPI), email addresses, etc.

The Value of Data Integration

No single agency has all the necessary data perfectly populated, and no single agency can identify all an individual's various touch points across diverse state agency programs and over time. Data integration consolidates various data sources, reducing the number of data silos and providing end users with a more comprehensive and unified view of the populations served. In addition, certain data elements present in one agency's data stores (such as social and demographic data) can become available for broader research purposes.

Other benefits of integrating data include:

- A better understanding of the whole person and their various touchpoints.
- The ability to fill in data gaps across datasets.
- Enabling comprehensive performance analytics and program evaluations across state agencies.
- Promotes data use and collaboration.
- Drives operational and technical efficiencies using a single interface and a common set of analytic tools to access the data.

IDS Program Examples: State Health and Human Services Agencies

Rhode Island Data Ecosystem

Established in 2017, the Rhode Island Data Ecosystem is an integrated data system that links multiple agencies' data at the person and family level. Through carefully governed, permissioned access to de-identified data, Ecosystem users conduct research projects that

explore how state residents are served across diverse state agency programs. Ecosystem priorities include developing tailored analytic files for efficient sharing of data among agencies; forming interagency analytic teams to identify cross-agency programs aligned with organizational priorities; and establishing performance analytics across agencies.

Governance and Roles: The Ecosystem is governed by four main "groups" or "roles" of stakeholders.

Table 1- RI Ecosystem Governance

Name	Meeting Frequency	Members	Purpose / Role
Executive Board ("Board")	Quarterly	Directors are data owners and sign Data Sharing Agreements (DSA) and Inter-Agency Memorandums of Understanding (I-MOUs)	Directors of agencies or organizations that submit data and other significant stakeholders (such as statewide IT agencies and Governor's office). The legal team advise directors but do not sign documents.
Data Stewards ("Stewards")	Monthly	One person per data set who has the expertise and authority to review and approve data use.	Sign project approvals and materials before publication. Advised by agency and community Subject Matter Experts (SMEs) as needed.
Project Advisors	As needed	State and community SMEs who guide projects from inception through dissemination.	Influence and advise but do not veto or approve.
Ecosystem Team	2X/Week	Staff	Implement direction of Board and Stewards; propose, support, and complete projects; maintain operations and the data asset; frame decisions for the Board and Stewards. Advise and guide the Board and Stewards.

Contributing Data Sources: Table 1 lists the datasets integrated into the Ecosystem. Nine state agencies and non-profit organizations, including the RI Coalition to End Homelessness, contribute and periodically update data sets.

Table 2: RI Ecosystem Integrated Data Sets

Agency	Fully Integrated Data Sets
Medicaid	Medical and pharmacy claims; enrollment
	Early intervention (EI) evaluation and encounters data
Department of Children, Youth, and Families (DCYF)	Case management (RICHST)
Department of Human Services (DHS)	TANF, SNAP, CCAP, SSI, & SSDI program eligibility (BRIDGES)
Department of Labor and Training (DLT)	Wages, employer, and industry records
	Temporary Disability Insurance (TDI) claims
	Temporary Caregiver Insurance (TCI) claims
	Unemployment Insurance (UI) and Pandemic Unemployment Assistance (PUA) claims
	Workers Comp data
	Real Pathways, Real Skills, and Real Jobs program data

Agency	Fully Integrated Data Sets
Department of Health (DOH)	Child screening, immunization, outreach program referrals (KIDSNET)
	Birth and death records (VITALS)
	Race and ethnicity data (VITALS)
	COVID testing, cases, and vaccination data
RI Coalition to End Homelessness (RICEH)	HMIS housing insecurity and homelessness data
All-Payer Claims Database (APCD)	Medical and pharmacy claims; enrollment *APCD data is not linkable to other Ecosystem datasets
Agency	**Newly Integrated Datasets**
Department of Corrections (DOC)	Incarceration, trial, and probation data
Unite Us	Social determinants of health data
Agency	**Limited Use Datasets** *Data is stored separately and can only be used for specific pre-approved projects
Department of Human Services (DHS)	Long-term Services & Supports (LTSS) application and billing data
Department of Health DOH	Prescription Drug Monitoring Program (PDMP) data
Department of Education (DE)	Education data

The program is funded via a Centers for Medicare & Medicaid Services (CMS) Implementation Advanced Planning Document (IAPD) enhanced match for a module within the state's Medicaid Enterprise System.

Hawaii Analytics Program

The Hawaii Analytics Program (HAP) is currently in development with a projected "go-live" in 2024. Led by Med-QUEST (Hawaii Medicaid), the University of Hawaii (UH), a public institution, will oversee this integrated data system housed within the Hawaii Medicaid Enterprise Environment. The vision for HAP, as outlined by Med-QUEST, is to streamline data standardization and alignment across various sources, bolster analytic and reporting capabilities, generate dashboards and interactive reports, and foster data sharing among Hawaii state agencies. The program is funded via a CMS IAPD enhanced match.

Governance and Roles: Multiple groups oversee, coordinate, and support the HAP activities.

Table 3: HI Analytics Program Governance

Name	Meeting Frequency	Members	Purpose / Role
Health Data Center Steering Committee	Quarterly	Directors or designees of high-level State Agencies and University of Hawaii (UH)	Provide guidance in alignment with state priorities. Review and approve policies.
Core Team	Weekly	Med-QUEST Health Analytics Office, UH, project management vendor	Review and finalize policies and framework. Provide day-to-day direction and project oversight.

Name	Meeting Frequency	Members	Purpose / Role
Data Review Board	Monthly	Data stewards and external subject matter experts (SMEs)	Make recommendations regarding the appropriateness of data requests
Data Governance Workgroup	Monthly	Core team members and data stewards	Raise data-source specific concerns. Provide input on data access policies, processes, and documents.
Data Stewards	As needed	Representatives of contributing data sources with healthcare data analysis experience	Recommend and guide process for data governance, quality, and compliance
External SMEs	As needed	State and community subject matter experts	Provide ad-hoc subject matter expertise. Influence and advise.

Contributing Data Sources: Data will be integrated into the HAP in a phased approach.

Table 4: HAP Data Sources

Agency	Fully Integrated Data Sets (At the Person-Level)
Medicaid	Adjudicated Claims and Encounters
	Beneficiary and Provider Data
	T-MSIS
	Enrollment and Eligibility
All-Payer Claims Database (APCD)	State Employee Enrollment, Claims (medical, pharmacy, dental), and Provider Data
CMS	Medicare FFS Data
Hawaii Level of Care Evaluation (HILOC)	Level of Care
Agency	To Be Integrated in Future
Managed Care Organization (MCO) Plans	Member and Provider Level Data
Department of Human Services (DHS)	Homeless Management Information System
Department of Health (DOH)	Child and Adolescent Mental Health Division Data
	Adult Mental Health Division Data
	Alcohol and Drug Abuse Division Data
	Hawaii Immunization Registry Data
Health Information Exchange (HIE)	Admissions, Discharge, Transfer (ADT) Data – Medicaid Beneficiaries
	Clinical Lab Data
	Provider Registry
Department of Human Services (DHS)	Child Welfare
	SNAP, TANF/TAONF
Department of Health (DOH)	Vital Records Data
Medicaid	Inmate Claims Data
TBD	Social Determinants of Health Data

Delaware Integrated Data System

The Delaware Integrated Data System (DIDS) is a Governor-initiated program to connect disparate state agency data sources and provide more comprehensive information to the Family Services Cabinet Council (FSCC) to inform service improvement for children and families in Delaware.

The Delaware Dept. of Technology and Infrastructure (DTI) is currently developing the DIDS, with policy support from the Office of the Governor, which staffs the FSCC. The state is taking advantage of the existing DTI technology stack in developing the DIDS. The program is funded by a CMS IAPD enhanced match.

Governance and Roles: DIDS is a cross-agency effort governed by representatives from the following offices:

- The Office of the Governor members include FSCC staff and other program appointees whose role is to negotiate use cases and data-sharing agreements, develop standard processes, and provide guidance in alignment with Governor's priorities.
- The DIDS Technical Team (DTI) includes technical leads (i.e., DTI staff and contracted technical support). Their role is to execute data governance documents and ingest, process, and integrate the data and to develop analytic products.
- The Medicaid technical team and its subject matter expert support contractor, business analysts and data scientists, and a local marketing/communications contractor.

Contributing Data Sources: Delaware is still in the planning phase, developing data sharing agreements. It is anticipated that the following agencies will be contributing data to DIDS.

- Dept of Agriculture
- Dept of Corrections
- Dept of Education
- Department of Labor
- Dept of Services for Children, Youth, and their Families
- Dept of Safety and Homeland Security
- Dept of Health and Social Services
- Dept of Technology and Information
- State Housing Authority

IDS Examples: Health Workforce Analytics

Washington Workforce System

The Washington Workforce Training and Education Coordinating Board is a governor-appointed partnership established in 1991 by the state legislature. The Board is responsible for coordinating efforts to enhance workforce development and education statewide. Its mission focuses on designing and implementing programs to provide training, education, and support services, in alignment with the goal of fostering a skilled workforce for Washington's economy. The Board evaluates the state's workforce system, advocates for continuous improvement and

innovation, supports career-connected learning under the Carl D. Perkins Act, regulates private career schools, and oversees veterans' programs. Historical data from 2017 is available in dashboards for many of the programs. Funding primarily comes from the Federal Workforce Innovation and Opportunity Act and the Wagner-Peyser Act.

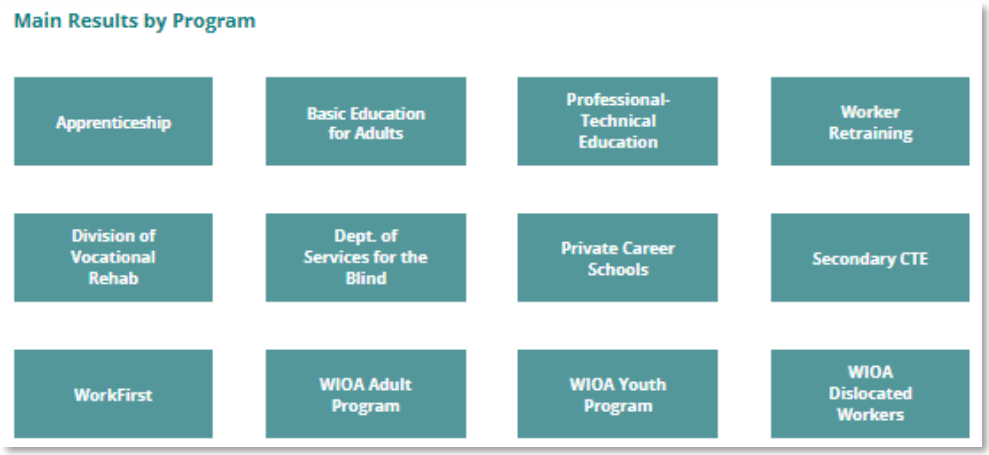
Governance and Roles: The Board consists of 9 voting members representing business, labor, and state agencies, along with non-voting members comprised of Board co-chairs and participating officials. Board members serve in an advisory capacity to both the governor and legislature, offering guidance on matters related to the state's workforce development system. Their responsibilities encompass planning, coordination, evaluation, monitoring, and policy analysis for the entire state's workforce development system.

Contributing Data Sources: 8 agencies contribute to the Washington Workforce System:

- Employment Security Department (ESD)
- Division of Vocational Rehabilitation (DVR)
- Department of Services for the Blind (DSB)
- Department of Social and Health Services (DSHS)
- Office of the Superintendent of Public Instruction (OSPI)
- State Board for Community and Technical Colleges (SBCTC)
- Department of Labor and Industries (LNI)
- Education Research and Data Center (ERDC)

Data Products: One of the highlights of WA's workforce system is its ability to track results by each of the 12 workforce training programs. The WA HWDC Results by Training Program dashboard yields a breakdown of the state's core indicator results along with the employment rates and annual earnings over time.

Figure 2 - WA Workforce Training Results

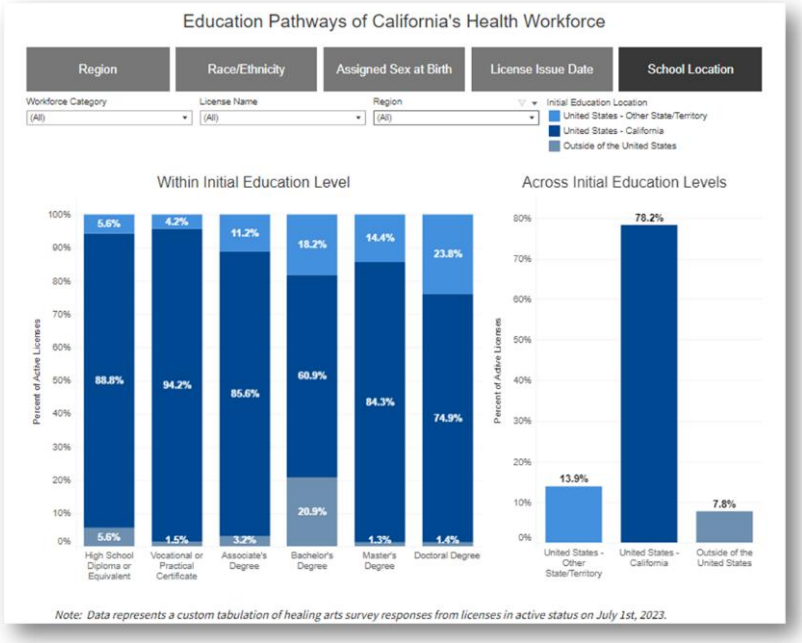


Additionally, there are many publicly available dashboards providing information on workforce demographics (e.g., age, sex, race and ethnicity, and prior education), employment outcomes across workforce programs (e.g., apprenticeships, private career schools, Wagner-Peyser Act), wages, workforce program information across industries, etc.

California's Health Workforce Research Data Center

The Health Workforce Research Data Center (HWRDC), operated by the California Department of Health Care Access and Information (HCAI), was established in July 2021 through Assembly Bill 133. Its primary purpose is to offer comprehensive data and analysis on workforce shortage, equity, and distribution within the health sector, aiding in the formulation of informed state policies. Functioning as the central repository for health workforce and health profession education data in California, it serves as a vital resource to inform and guide state policy decisions related to healthcare workforce issues.

Figure 3 - CA's Health Workforce Education Pathways



Contributing Data Sources: The California Health and Human Services Agency (CAL HHS) and the Department of Consumer Affairs are current data contributors to the HWRDC. The HWRDC also aims to incorporate other types of data into their system to allow for further comprehension and understanding of the healthcare workforce, including their qualifications, training pathways, and areas of expertise.

Table 5: CA HWRDC Data Sources

Current Data Sources
Healthcare Workforce Geography Crosswalk
Physicians by Specialty and Patient Care Hours
Registered Nurse Shortage Areas in California
Primary Care Shortage Areas in California
Health Professional Shortage Areas in California
Physician Survey – Licensee Responses by Address of Record ZIP Code
Data from licensed health care professionals at time of licensure renewal
Planned Data Sources
Health Workforce Licensure Surveys: These surveys gather information about the licensing status of healthcare professionals, providing insights into their qualifications and regulatory standing.
Education Program Data for Health Professions: This includes data from educational institutions offering programs related to health professions, such as medical schools, nursing programs, and allied health programs. This data can shed light on the number of students enrolled, graduation rates, specialties pursued, etc.
Health Workforce Data for Certified Professions: This involves data regarding professionals who have obtained certifications in various healthcare fields. This could encompass certifications for specific skills or specialties within the healthcare industry.

Data Products: The HCAI HWRDC currently publishes several dashboards with additional

views expected per the HWRDC roadmap. The initial dashboards are focused on workforce demographics and education and include:

- Race & Ethnicity of California's Health Workforce
- Languages spoken by California's Health Workforce
- California's Health Workforce Education Pathways
- California's Registered Nursing Education Pathways

Virginia's Healthcare Workforce Data Center

The VA Department of Health Professions' (DHP) Healthcare Workforce Data Center (HWDC) was established in 2008 with a mission to improve data collection and measurement of Virginia's healthcare workforce through regular assessment of workforce supply and demand issues among the 80 professions and 350,000 practitioners licensed in Virginia by the DHP.

Governance and Roles: The VA DHP HWDC is overseen by advisory committees and workgroups along with an advisory council of 17 key stakeholders representing over 100health professionals and data network members. The committees and workgroups provide advice and expertise with development of the data center, workforce surveys, and reporting information on the website.

DHP Healthcare Workforce Advisory Committees

- ASLP (*Audiology & Speech-Language Pathology*)
- Behavioral Science
- Dentistry
- Healthcare
- Long Term Care
- Nursing
- Pharmacy
- Physicians

Workgroups

- CNA Workforce
- Nurse Practitioner Workforce
- Physician Assistant Workforce

Contributing Data Sources: VA DHP HWDC incorporates various contributing data sources. Among these are voluntary surveys integrated into the department's online application and renewal processes, allowing for the collection of data regarding the qualifications and demographics of healthcare professionals. As each renewal cycle concludes, survey reports are compiled for each profession and then released, providing timely and comprehensive insights into the workforce landscape.

Data Products: The collected data plays a vital role in data products produced by the VA DHP HWDC. The data products include profession-specific reports, data visualizations via public dashboards, and several Virginia Healthcare Workforce Briefs and occasional papers and reports regarding retirement intentions by profession and other topics.

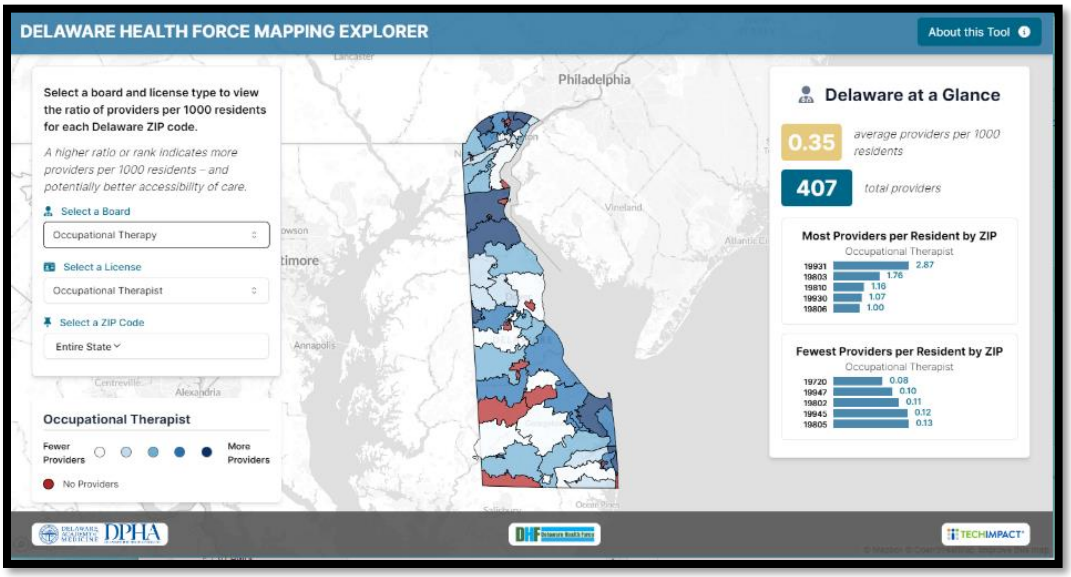
Other HWDCs

There are Other HWDCs not fully explored in this document. Some are listed below:

- The National Database of Nursing Quality Indicators® (NDNQI®) The only national nursing database that provides quarterly and annual reporting of structure, process, and outcome indicators to evaluate nursing care at the unit level.
- Texas Health Professions Resource Center (HPRC) Analyzes healthcare workforce data, reports findings, designates practice sites, and offers resources for collaborative practice through a clearinghouse program.

- Health Workforce Data, Tools, and Dashboards Accessible data on Health Workforce programs, including user-friendly dashboards for loan and scholarship awards and downloadable information on grant funding for health professions training programs.
- Health Professions Tracking Service (HPTS) A data system that tracks education, training, and employment of various healthcare professionals over time.
  Delaware Health Force Mapping Explorer The Delaware Health Force Mapping Explorer provides data on healthcare service accessibility across Delaware, presented at the ZIP Code Tabulation Area (ZCTA) level and in provider-to-resident ratios. Developed using data from the Delaware Professional Regulation Online Services (DELPROS) and the National Plan & Provider Enumeration System (NPPES) datasets, refined for analysis, and complemented by population data from the U.S. Census Bureau's American Community Survey 2022 5-Year Estimates, it serves as a foundational resource for assessing healthcare workforce accessibility.

Figure 4 - DE Health Force Mapping Explorer



Lessons Learned:

- Develop compelling use cases and identify resource needs while planning for sustainability.
- Explore strategic partnerships and conference opportunities to broaden reach.
- Dedicate time to writing legal documents to establish trust with data owners.
- Expand data and reporting capabilities gradually, phasing in use cases.
- Develop and test customized person matching and leverage an enterprise provider directory.
- Gain expertise in federal funding and progress reporting.
- Assess resource needs and consider the advantages and disadvantages of hiring versus outsourcing.
- Evaluate vendor and solution requirements and preferences.
- Understand industry standards.

CHAPTER 2: DATA GOVERNANCE

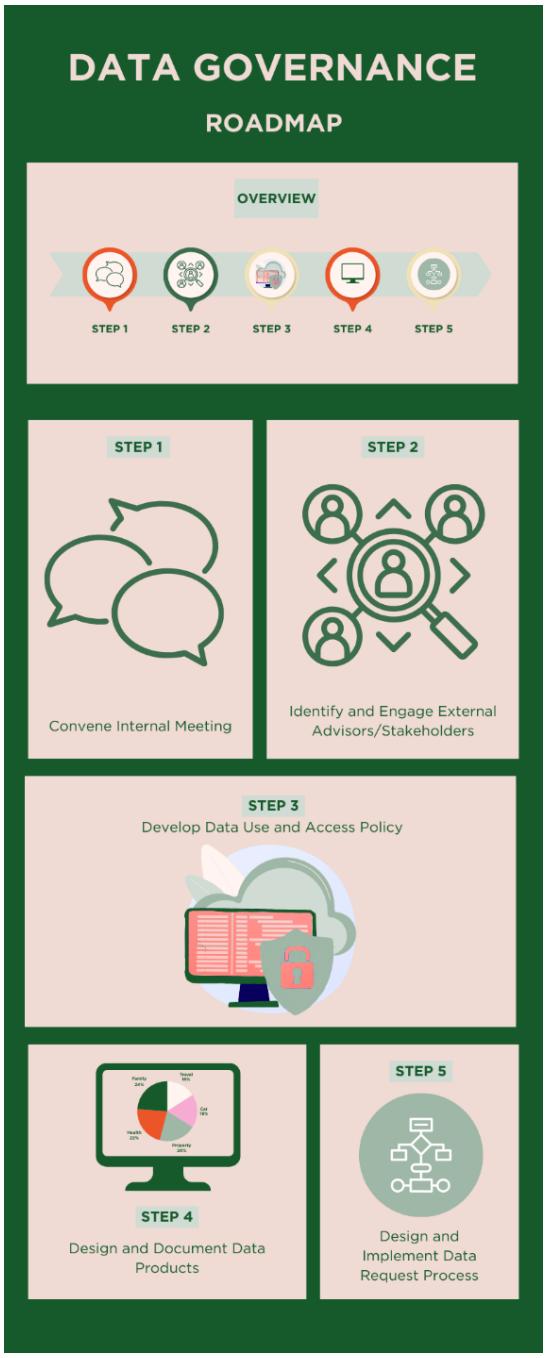
Introduction

Data Governance is a set of policies, procedures, and related documentation that guides how people will manage data to create actionable information and maximize value. Broadly, data governance manages who will access what data, when, and under what circumstances. Applied to an Integrated Data System (IDS), data governance is the practice of knowing where the data came from, how it will be utilized, and how it will be adequately protected. Topics include identifying the core operations team; stakeholder engagement through committee participation and well-defined roles; governance issues related to data use, access, and release, including pathways for accessing different types of data products; and public reporting. This chapter offers a step-by-step outline of the processes.

Key Takeaways:

- Identify teams, leaders, and stakeholders that will support both the vision and the operations of the IDS, including modular components (e.g., master person index, provider directory).
- Define the data governance roles and responsibilities.
- Build an effective data governance team by providing orientation, education, and training.
- Communicate effectively and frequently to build trust and a collaborative environment.
- Establish procedures to safeguard the data throughout the process, from data collection, integration, storage, and release.

Figure 5 - Data Governance Roadmap

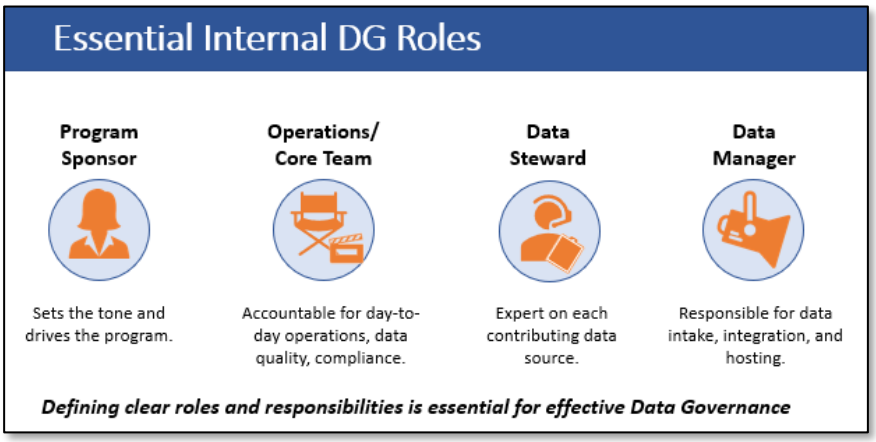


Data Governance Roles

Internal Team

IDS programs begin by identifying the "Internal Team" that will support the day-to-day operations of the database. Typically, there are four internal data governance roles that need to be defined and filled as part of this internal team. These are:

Figure 6- Internal Data Governance Roles



Program Sponsor: Leader of the entity or organization that will collect and manage the data. This individual is responsible for:

- Creating value through the IDS consistent with the mission and program goals.
- Managing the relationships between the "host" agency, internal staff, and stakeholders
- Accountable to entities with statutory and/or regulatory authority for the IDS

Operations: Individual(s) responsible for making the decisions (with stakeholder input) related to data governance. Responsible for:

- Ensuring data quality, accuracy and completeness, appropriate data access, and allowable uses.
- Leading stakeholder engagement processes to inform data governance policies and procedures.
- Performing data governance oversight throughout the program and data lifecycle.

Data Steward: Subject matter experts from each data source collected by the IDS. These individuals have access to necessary data dictionaries and understand the data quality processes for the source data. The data stewards are responsible for:

- Providing information on what data can be shared with the IDS, how frequently, and in what format.
- Working with the operations team to answer questions related to specific data fields and definitions.
- Resolving data quality issues and questions.
- Participating in the data request and approval process.

Data Manager: Individual(s) responsible for the technical aspects of data governance, including:

- Overseeing day-to-day IT operations (e.g., data intake, integration, and hosting).
- Protecting data privacy and security and preventing inappropriate access/use.
- Maintaining a secure storage environment throughout the data lifecycle.

Stakeholder Engagement: Getting Buy-In to Share Data

Once an internal team is identified, IDS programs often turn to identifying additional stakeholders that can support, amplify, and expand the database in meaningful ways. Specifically, successful IDS programs have data governance frameworks that incorporate the following elements related to stakeholder engagement:

- Identify key stakeholders and assign them clear roles and responsibilities.
- Collaborate closely with stakeholders to understand their data needs and concerns early on.
- Leverage subject matter experts in developing the policies and procedures that will govern the IDS.
- Provide all stakeholders with orientation, education, and training to promote participation, acceptance, and buy-in.
- Work with stakeholders to develop flexible allowable use cases for the IDS data, which are consistent with the enabling legislation and regulations, to allow for the broadest downstream applications.

IDS programs begin by identifying categories of stakeholders to engage in the discussions and efforts itemized above and then identify individuals with the availability and interest to serve in each category. At a minimum, these stakeholder categories generally include the following:

Figure 7 - Categories of Stakeholders to Engage

- **Program Champion:** An executive-level leader or advocate for the program who can promote its advancement (*can be the Program Sponsor of the internal team).
- **Data Owner:** Individuals who will host, operate and/or maintain the database.
- **Data User:** Individuals who will use the data within the database.
- **Internal Expert:** Individuals with expertise on specific topics related to database operations, data governance, data architecture, downstream use, legislation, etc.

Stakeholder Committees

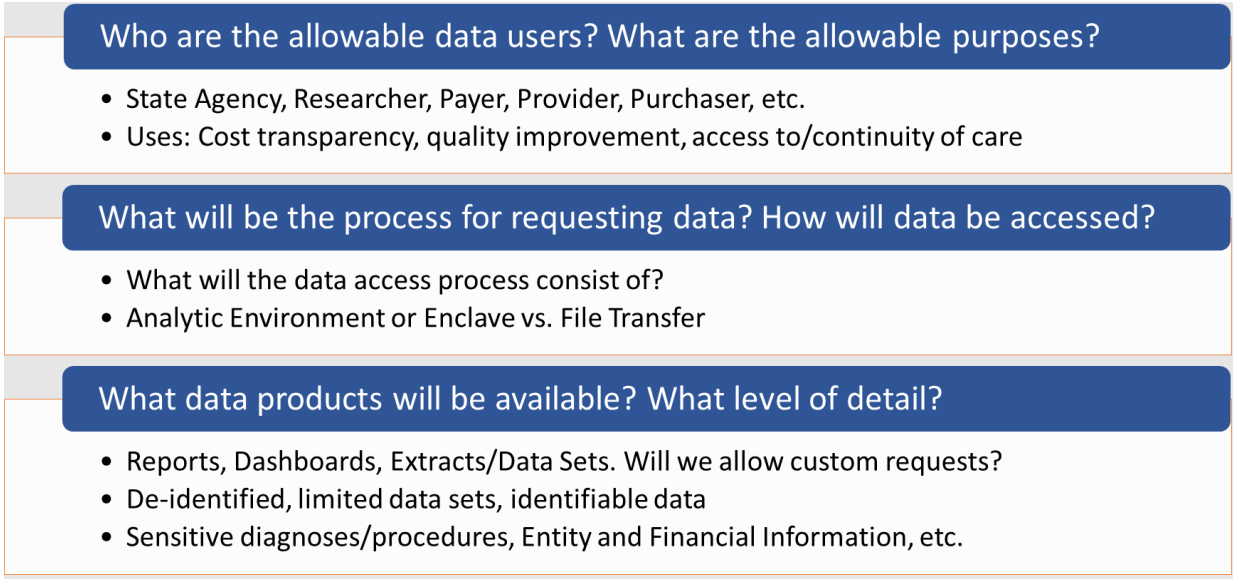
IDS programs should convene their stakeholders early, frequently, and continuously. Many seek internal and external input through two different groups with distinct roles and functions. As noted below, advisory and data access committees facilitate stakeholder representation and engagement, creating a forum to obtain input from a range of experts and perspectives.

Both committees can advise the Internal Team. State agencies typically retain authority for final decisions on accepting a committee's recommendation or approving an access request.

Committee membership should be carefully considered. Sometimes the enabling IDS legislation or regulations specify who needs to serve on the committee. For example, specific stakeholder groups must be represented to ensure that diverse perspectives are shared, and common needs and concerns are represented. Other times, committee membership is solely the decision of the Internal Team. Decisions on who should serve on committees should consider members' availability and willingness to participate, especially when quorum needs to be met for public meetings; goals of multi-disciplinary stakeholder involvement for buy-in and transparency; and an eye towards efficiency (e.g., odd numbers can serve as tiebreakers).

Data Governance: Use and Access

A major set of data governance decisions that both the internal team and stakeholders will need to weigh in on are questions related to data use and access within the IDS. These policy and procedural decisions can be grouped together into the following three question sets:



Allowable Use Cases

The term "use case" is used to describe scenarios in which data from the IDS could potentially be used, with the objective of identifying important requirements that should be proactively considered at the design stage. Identifying and evaluating the merit and appropriateness of various use cases at the outset of an IDS, helps shape operational features of the program such as the type of data sources and data elements that will be collected, data management processes, data access strategies, and identification of analytic and reporting priorities.

Though there are myriad IDS use cases, it is important to meet with the internal team and stakeholders to identify the ones they find compelling, feasible, and consistent with the intent of the IDS. When identifying use cases, the following framework can be helpful:

1. **Key Question**: What question is the user trying to answer?
2. **Overview**: What information will need to be available to the user to answer this question? For example, statewide summary statistics on the number of professional licenses granted each year, by healthcare profession.
3. **Audiences, Primary and Secondary**: What stakeholder audience will likely be interested in this use case? Which ones are the target, or primary, users, and which ones indirect, or secondary?
4. **Output Examples**: What would be the preferred format of the analytic product for this use case? Examples include raw data available to requestor, analytic report, or interactive dashboards.
5. **Policy and Business Value**: What is the desired outcome or value proposition to the user? Examples include identifying healthcare access shortages, providing information on market share, or supporting new training programs.
6. **Other State Examples**: What other states, if any, have pursued a similar use case example?

Next, the internal team and stakeholders should consider whether some use cases are so misaligned with the purpose of IDS that they should be prohibited. For example, can the IDS be used by an external, commercial entity such as a pharmaceutical company that wants to use the information to target specific provider groups in a different state?

Once salient and prohibited use cases are identified, a successful IDS develops language to enable the *broadest* application of allowable use cases. These allowable use cases must align with the statutory purpose of the IDS and comply with any existing state and federal legislative and regulatory requirements. Once developed, the list of allowable use cases will become the first criteria on which to evaluate applications/requests for data from the IDS.

Permitted Data User Categories

In addition to determining the allowable use cases for the IDS data, new programs should consider *who* will be allowed to access the data within the IDS. This will become the second criteria for evaluating requests for data and will inform the development of the data architecture and analytic products. Different access provisions apply to internal requesters compared to external requesters:

- **Internal Requestors:** Users from within the walls of the organization/agency that maintains the database or contributes data (or their contractors/proxies).
- **External Requestors:** Users from outside the walls of the organization/agency that maintains the database or contributes data.

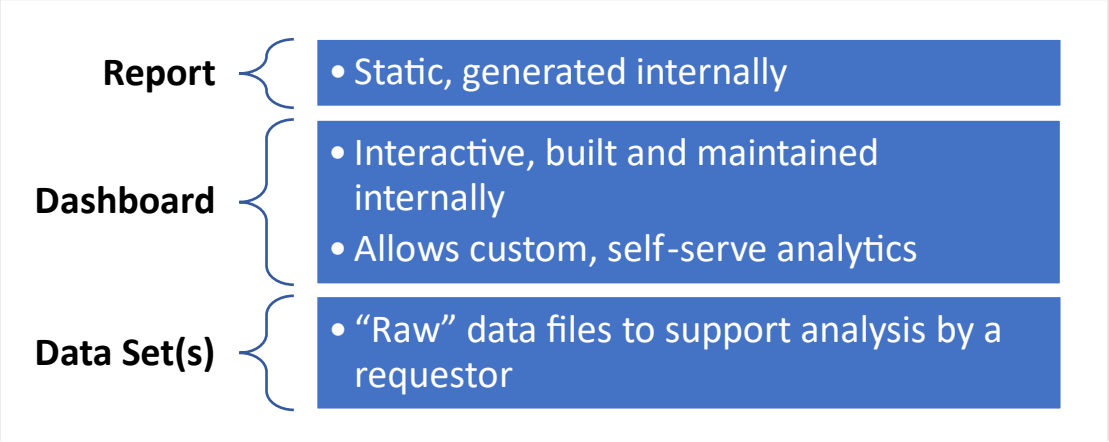
Stakeholders and Internal Teams should consider whether the IDS will allow both internal and external requestors, or if just internal requestors will be permitted to access data. As a reminder, the allowable use cases would still apply in all instances. For example, even an internal requestor would need to be requesting data for an allowable use case. If external requestors are permitted, what is their obligation, if any, to share their findings or analytic output with members of the Internal Team? It should be noted that choosing to review users' outputs places a burden on state agency staff and creates an appearance of state endorsement (or rejection) of findings and a statement noting otherwise should be provided.

Some IDS programs start off with just allowing internal requestors for the first few years of a program to build trust and fully develop all policies, and then move to allowing external requestors later (e.g., the Rhode Island Ecosystem). The question of whether to allow both internal and external requestors is also often contingent on the availability of funds to sustain the IDS, as most revenue-generating products are developed with external requestors in mind.

Data Products: Types and Level of Detail

Data products are the types of data outputs that will be available to allowable users for the available use cases. The most common data products are static reports, interactive dashboards, and raw data sets. IDS teams should consider which types of data products will be available for request, and whether they will all be available at the outset or whether they will be phased in through a tiered approach.

Figure 8 - Common Data Products



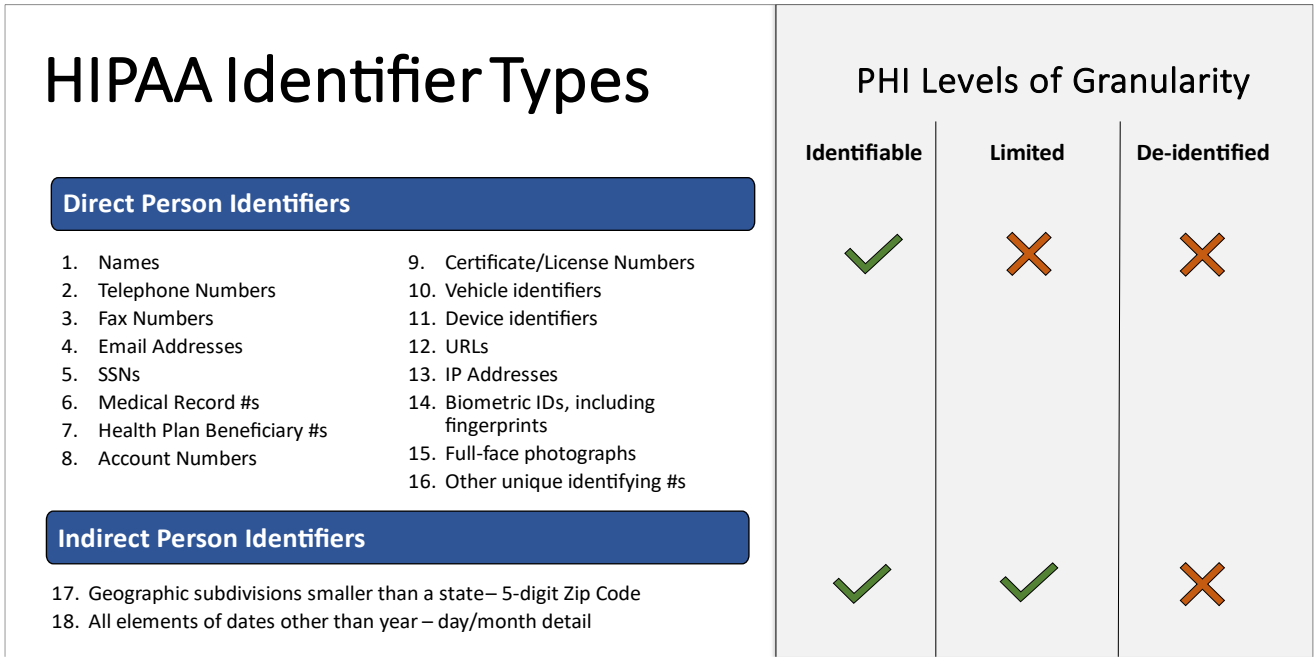
Level of Detail: Data Set Sensitivity

In addition to determining which data products will be available, IDS teams need to consider what level of detail will be available within each of the data product types. The combination of these considerations (i.e., type and detail) provides the necessary framework for the IDS's analytic plan. For example, many IDS teams managing health information use the HIPAA Privacy Rule[2] , and the associated definition of identifiers (see Figure 9 below), to delineate the levels of detail. A similar structure of data products may be implemented for workforce data.

1. **Identifiable Data:** Data that includes direct HIPAA identifiers.
2. **Limited Data Set:** Data that includes indirect HIPAA identifiers only (e.g., no direct identifiers).
3. **De-Identified Data:** Data that includes neither direct nor indirect HIPAA identifiers.

[2] https://www.hhs.gov/hipaa/for-professionals/privacy/index.html

Figure 9 - HIPAA Identifiers



Many states develop augmented data request processes depending on the level of detail available and being requested. For example, requests for de-identified data from internal requestors may be eligible for an expedited review that doesn't require application review by the Data Access Committee. IDS teams should meet with stakeholders to consider all these data access and release issues in tandem, to develop the most appropriate pathways and offerings for their programs.

Data Request Process

Prospective users of data that is not otherwise publicly available (e.g., not available in a public-facing dashboard) must complete a data request application. The data request application ensures that:

- the proposed use of the data is consistent with established allowable uses for the IDS,
- the data requestor is in a permitted user category, and
- the requested data is suitable for the research question.

Once an application is submitted, an Internal Team member reviews the application and, as appropriate, schedules follow-up discussions with the requestor. The final request documentation is presented to the Data Access Committee for consideration.

Once an application is approved by the Data Access Committee, a Data Use Agreement (DUA) is typically signed. The approved application is incorporated into the DUA by reference and as

an exhibit. Common components of both documents are included in Figure 10 - Data Request & Data Use Agreement Components10 below.

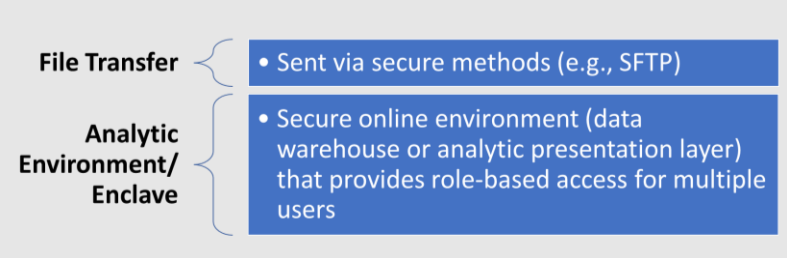
Figure 10 - Data Request & Data Use Agreement Components

Data Request Application:	Data Use Agreement (DUA)
• Project purpose and intended use of the data • Name, title, affiliation, and qualifications of requestor, including experience working with similar data • Type of data/specific elements requested, including justification as needed • Plans to link the data to another data source • Methods of analysis and expected results/outputs • Information on how the results will be used, shared, or disseminated	• Requires adherence to outlined privacy and security requirements • Data will be used only for purposes listed in the approved application • Data will be accessed only by the requestor and/or approved members of the project team • Data recipient will not use the data to re-identify individuals by linking to other data sets • Data user will certify that the data has been destroyed after its use/project end

Data Access Methods

There are generally two ways that approved data can be shared with the approved user – a file transfer or direct access via an analytic or "enclave" environment. The former data access method is significantly easier to establish and cheaper to maintain and may be required by the IT standards of the host organization. The latter is often desirable for end users and administrative staff but can be considerably more costly. The IDS teams need to determine what their budget, analytic architecture, and data governance rules allow.

Figure 11 - Data Access Methods



Summary: Data Governance Decisions

Category	Description
Access Group	**Internal:** Within the walls of the agency that contributes that data or their proxy/contractors. **External:** Outside the walls of the agency that contributes that data. What external users will be allowed?
Allowable Uses	What purposes can the data be used for?
Data Access Process	What will be the process of requesting data? What documents will be part of this process? Will there be an augmented/expedited process for certain types of requests?
Data Products	**Reports:** Can be produced and published as a standard or by custom request**.** **Dashboards:** Visualization layer available to end users. Data in the backend are fully processed. Users can interact with, filter, manipulate data and view results. Used with Standard Data only. **Data Sets:** Data file to be accessed by requestor. Can be standard data sets or custom data sets.
Data Set Sensitivity	What sensitive data is included in the data sets (e.g., Social Security Number, date of birth)? Will sensitive data be released? Or transformed (e.g., transforming date of birth to age in years or age ranges)?

CHAPTER 3: BEST PRACTICES IN DATA ARCHITECTURE

Introduction

Data architecture focuses on the way in which an organization's policies, standards, and rules are structured to fulfill an organization's business needs. In the case of IDSs, business needs often call for the integration of datasets. A successful integrated data architecture will include an assessment of existing infrastructure, the implementation of meaningful metadata, and the integration of data governance systems to consistently monitor the lifecycle and quality of data assets.

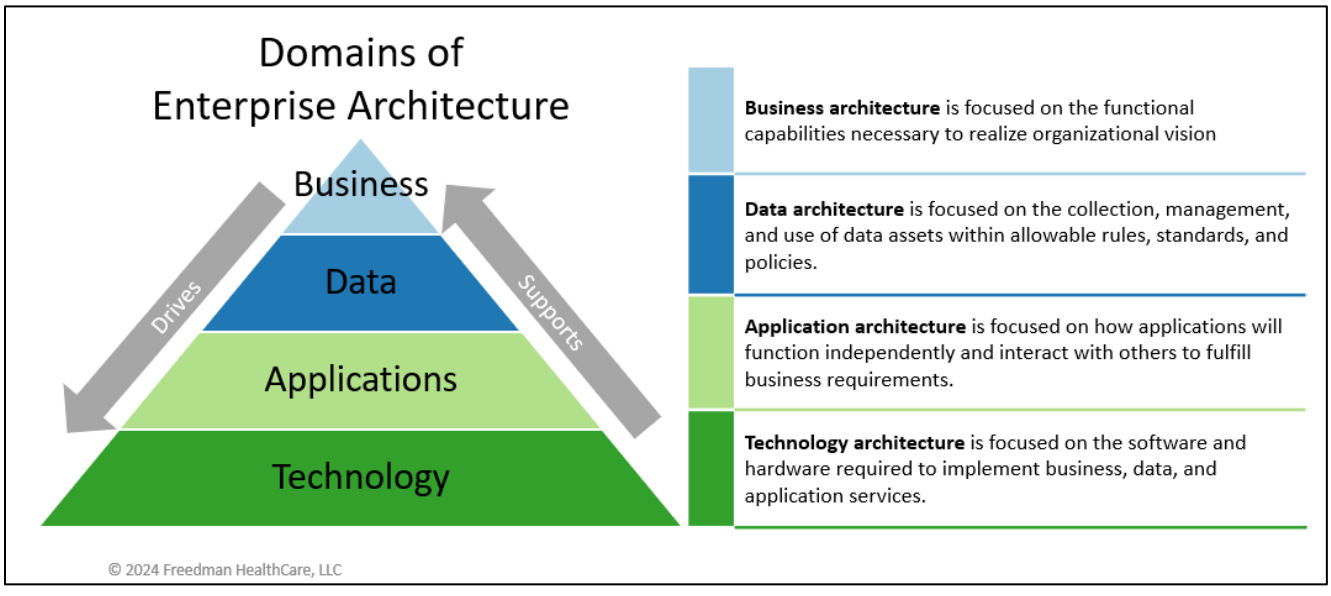
Key Takeaways:

- Business needs should drive technological decisions.
- Data security and protection are critical.
- Effective systems include dedicated resources for master data management, including metadata and identity management.
- Evaluating opportunities to extend or reuse existing architecture should be done carefully, considering regulatory compliance and modularity.
- An integrated data system is built on trust and the architecture should embed quality assurance and security throughout the data lifecycle.

Data Architecture Supports Business Needs

Information Technology (IT) incorporates an organization's overall Enterprise Architecture (EA), which includes the holistic structure and coordination designed to meet its business requirements. Business architecture drives the other layers of EA, which should be designed to support the organization's vision.

Figure 12 - Enterprise Architecture Domains



In the data architecture layer, the data considerations should always be considered in

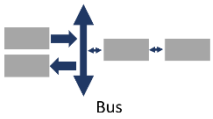
relationship to how important they are in fulfilling the business needs. Selecting appropriate software and technical infrastructure should be guided by an understanding of the required data and how it can be leveraged to meet business needs. IDSs, like other integrated data systems, will require sophisticated business rules, monitoring, and security requirements so that they may reliably apply appropriate data governance.

When integrating data, architecture should employ substantial metadata to help establish a source of truth, depending on the business case. Consistent review of these business rules should be integrated into quality monitoring processes, especially with the introduction of new datasets into the IDS.

By bringing data together, the organization may be able to glean a more comprehensive view of its business population, leading to more powerful and nuanced insights. Integration also provides a chance to efficiently apply cohesive security and privacy protection, implement consistent data treatment, and centralize business logic.

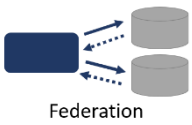
There are many different potential architectural models that may be implemented as part or in totality when data is being integrated, including:

Hub and spoke: A centralized "hub" collects data from various sources (or "spokes"). Data are subsequently distributed to integrated endpoints or systems. Once advantage of this mode is that control and management occur together, which may facilitate the integration of additional data assets.



Bus: The "bus" serves as a conduit for data, allowing movement among different systems. This type of architecture is especially helpful for real-time business needs, as transfer between integration points can be rapid.

Pipeline: Data pipelines systemically stage the ingestion, transformation and loading information to a dedicated endpoint, often a data warehouse. A pipeline can be useful for complex integration scenarios, as it easily accommodates multiple coordinated data streams.



Federation: Federated data systems allow data assets to remain distributed and provide a virtual view that brings data together as if they were resourced together. One advantage is that it improves security and performance by reducing the need to transmit data; however, the trade-off is that they are often complex to design and implement.

Design Considerations

After identifying the data needed to implement the business vision or need, planning should consider the volume, velocity, and variety of information, including assessment of any risks

associated with changes to them over time. The complexity of solutions is often increased with higher amounts of data, “real-time” transfers of data, and/or several, disparate sources of information. Many of these considerations can be addressed when considering the solution’s needs for interoperability, scalability, and modularity.

Interoperability: How easily the solution or data can work with others

Figure 13 - Interoperability

When discussing interoperability, we refer to how easily the solution or data can work with others. Interoperable systems, products, or data work well together with other systems, products, or data without special effort on the part of the user. In other words, they seamlessly exchange information.

Considerations for Selecting an Architecture

When considering the adoption of a data warehousing and business intelligence solution, it's essential to recognize the multitude of variables that influence the decision-making process. The selection of an architecture must align closely with the organization's unique needs and objectives.

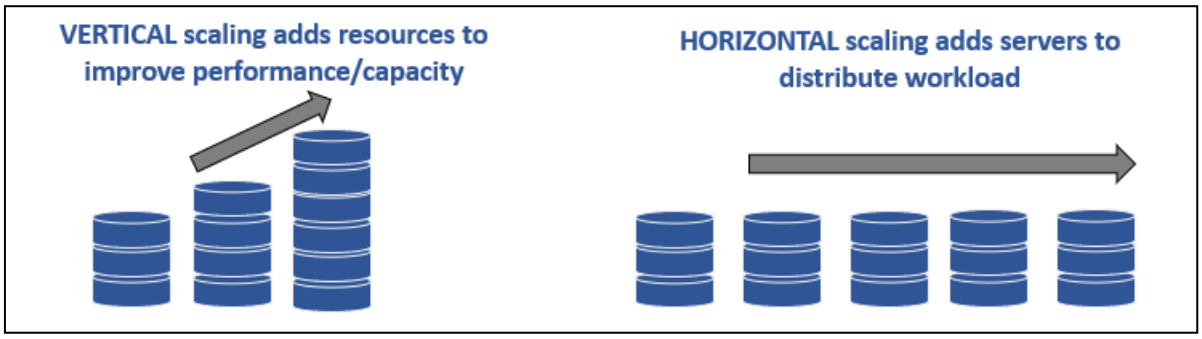
1. **Current and Future Needs:** The chosen architecture should not only address current requirements but also anticipate future needs. This forward-thinking approach ensures scalability and adaptability as the organization evolves.
2. **Cost and Time Investment:** It's crucial to acknowledge that implementing data solutions can be a significant investment both in terms of cost and time. Developing and deploying these solutions may require substantial financial resources and an extended development timeline.
3. **Value Proposition:** Despite the investment required, data solutions represent invaluable assets in facilitating data-driven decision-making processes within the organization. The insights derived from these solutions have the potential to drive strategic initiatives and foster organizational growth.
4. **Data Source Evaluation:** When assessing how interoperable a new data asset is with an existing IDS, it is important to consider how well-aligned it is with existing assets. Where there is overlap, it’s helpful to consider the strengths and limitations of each resource to determine which is more trustworthy in the hierarchy of the "source of truth."
5. **Automated Processes:** Additionally, interoperability may factor into additional data when considering how much manual configuration is required to integrate the data, as manual steps often introduce errors and other risks to data quality.
6. **Effective Metadata:** In IDSs, effective metadata is essential for making data work well together. It provides cross-data asset information, such as a master person index, which

allows the same person to have the same identifier across time and across data assets. Metadata also provides the mechanism to apply and track the appropriate treatment of data across its lifecycle, along with information such as the frequency of ingestion, subject matter experts, and/or stewards.

Scalability: how the solution can adapt to workload

Scalability refers to how well the solution can adapt to varying workloads. Like other complex IT systems, IDSs may experience fluctuations in demand. Effective systems should scale as needed to meet business requirements and maintain seamless performance and functionality.

Figure 14 – Vertical and Horizontal Scaling



Systems may be designed to scale *vertically* meaning that additional resources are added to improve capacity and performance. Others may scale *horizontally*, which means additional servers are added to distribute the workload. Ideally, an IDS's scalability will include requirements for how elastic it needs to be when scaling up to handle increases in workload and when scaling back down when demand recedes.

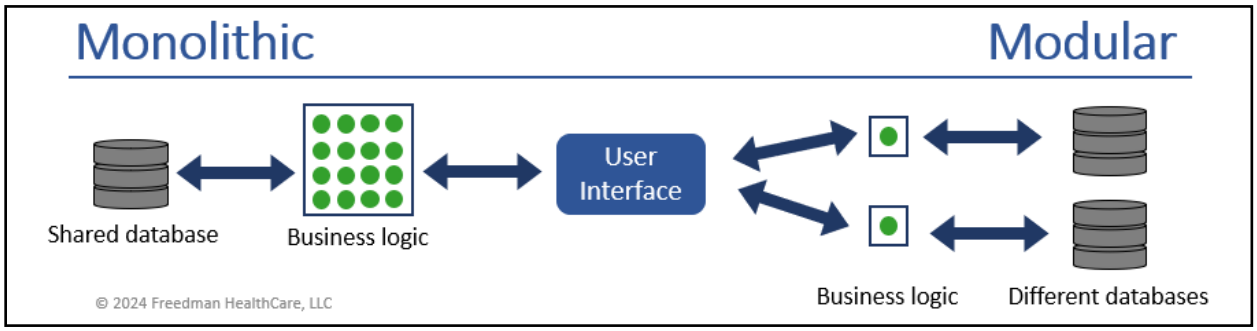
Scalability in data systems encompasses two primary capabilities: scaling up, scaling down, and the ability to scale to zero. The ability to scale up allows us to handle extreme loads temporarily, ensuring that the system can accommodate spikes in demand effectively. Some scalable systems can also scale to zero, meaning they shut down completely when not in use. Once the load spikes subside, it is essential to automatically remove excess capacity to optimize costs. An elastic system can dynamically scale in response to load fluctuations, ideally in an automated manner. This dynamic scalability ensures that the system can adapt seamlessly to changing workload demands, maximizing efficiency and resource utilization.

Modularity: how independently components of data models and infrastructure function

Modularity refers to how components of data models and infrastructure independently function. Modularity for systems and data infrastructure involves trade-offs. Modular systems offer small, discrete components that work together, enhancing the adaptability of the system and enabling the "plug and play" integration of new components as technology emerges. Additionally, modular systems are often less fragile, meaning that if a component fails, the system can continue to perform. However, excessive modularity can make debugging issues and enhancing systems much more complex to administer. In contrast, a monolithic system is one that encompasses all functionalities, whereas modular data architecture segments data/capabilities

and integrates them into the user interface.

Figure 15 - Monolithic vs. Modular Systems



Data Hosting and Infrastructure

IDS components may be hosted in the cloud, which are remote servers hosted by third parties and accessed via the internet, or on-premises, which are servers and other infrastructure managed locally.

Organizations should carefully consider their existing infrastructure and assess which option is most promising for reuse, as opposed to those that require new, dedicated infrastructure. Larger organizations tend to be better positioned to invest in on-premises solutions. Sometimes, governance or other security considerations necessitate an on-premises solution. In contrast, cloud-based solutions offer more flexibility for data sharing and provide sophisticated capabilities out-of-the-box.

Other trade-offs between cloud and on-premises hosting include the relative ease of data backup and recovery, scalability, and administration of data security that the cloud offers. Cloud solutions also tend to be more affordable and leverage cost models on a per-user or per-instance basis. Conversely, on-premises solutions offer the most control and responsiveness tailored to business needs. Figure 16 summarizes these considerations.

Figure 16 - Infrastructure: Cloud vs. On-Premises

	Cloud	On-premises	
Control		☑	Local infrastructure allows the most control
Cost	☑		Upfront expenses are substantially higher for on-premises solutions
Data backup	☑		Backup and disaster recovery are provided by the vendor
Performance	☑	☑	Triaging and development are dictated internally for on-prem solutions, but cloud-based resources offer superior elasticity
Scalability	☑		Cloud tools usually are priced based on utilization of resources (e.g., per user licensing)
Security	☑	☑	There is more potential exposure for risks associated with unauthorized data access or breach when it is transmitted, however cloud solutions may have more experienced, dedicated personnel

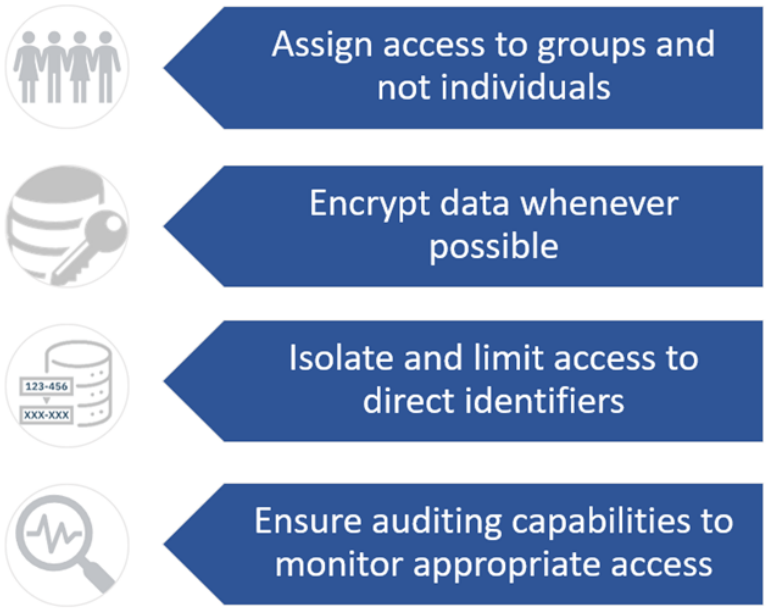
Some organizations opt for a hybrid approach, combining elements of both cloud and on-premises hosting to leverage the benefits of each. In a hybrid setup, certain components may be hosted in the cloud, while others remain on-premises. This approach offers flexibility, allowing organizations to balance factors such as data sensitivity, performance requirements, and cost considerations. By strategically allocating resources between cloud and on-premises environments, organizations can optimize their infrastructure to best suit their needs.

Architectural Considerations for Governance, Security, and Privacy

Data governance is an essential and complex topic within the context of data architecture. Governance encompasses the overall management of data, including its availability, usability, integrity, and security.

Effective governance entails documenting the lineage of data, which involves metadata tracking its lifecycle from the moment it enters the IDS to its current state. The traceability of data is also a key consideration, as it evaluates how well the data conforms to its intended accuracy, quality, and reliability.

Figure 17 - Principles for Secure Data Management



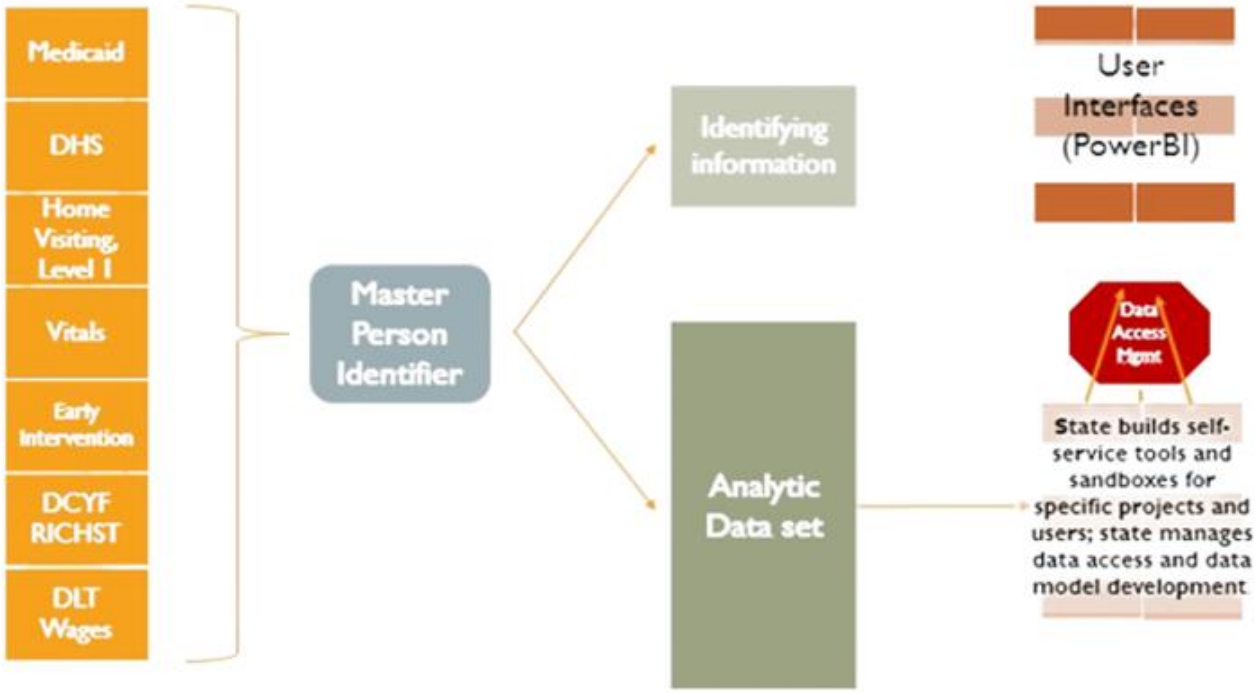
Governance is designed to safeguard data privacy by implementing security measures. Privacy means limiting data to its intended use and only disclosing what is necessary to help protect the individuals represented by the data. Security policies, protocols, and other standards are applied to help enforce appropriate use, access, and disclosure.

Example of Data Architecture in Other IDSs

RI Ecosystem

Figure 18 below provides an overview of the Rhode Island Ecosystem. On the left side are various datasets, each flowing through the Master Person Identifier before branching into identifying information and the Analytics Dataset, which is made accessible to other agencies. Subsequently, the state develops self-service tools tailored to specific projects and users, while also overseeing data access and model development.

Figure 18 - Rhode Island Ecosystem



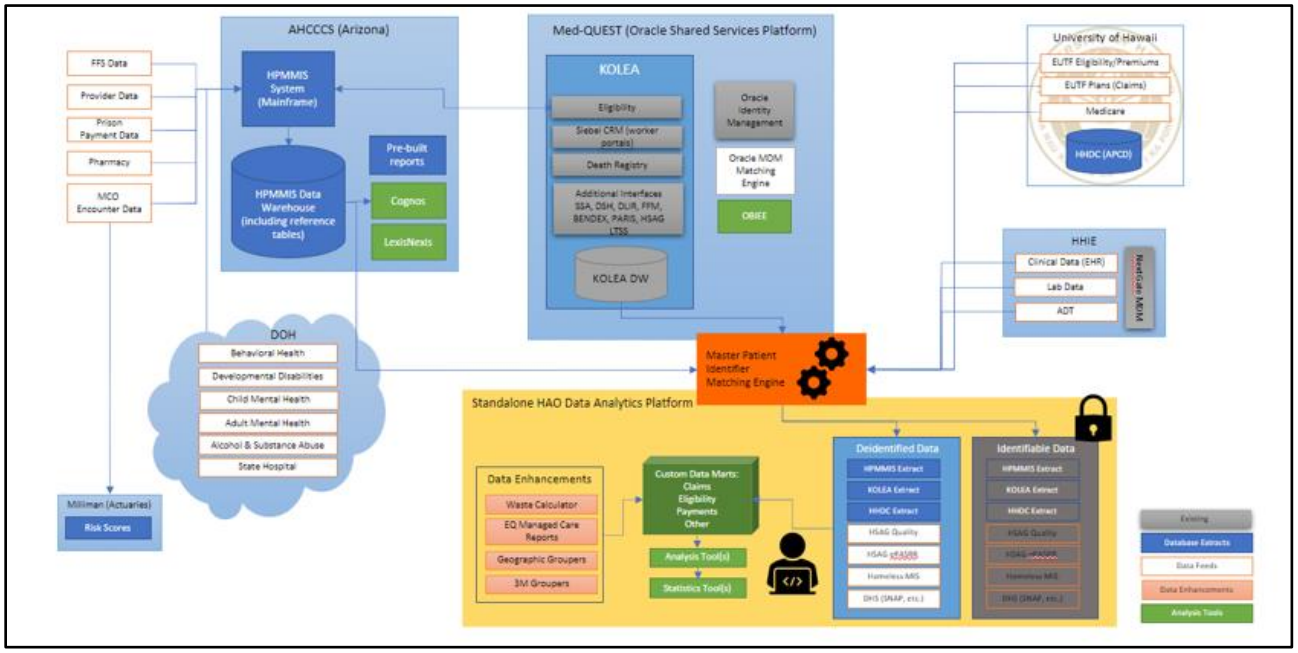
Hawaii Analytics Program

Hawaii's Data Environment comprises various datasets that flow into the Master Patient Identifier Matching engine. Subsequently, the data is transferred to a secure Standalone HAO Data Analytics Platform, where it undergoes de-identification processes.

In response to Hawaii's specific requirements, the decision was made to leverage their existing Oracle system within this data environment instead of embedding the HAO Data Analytics Platform. Several factors influenced this decision:

- **Independence:** The chosen environment can operate independently, providing Hawaii with flexibility in deployment, upgrades, and maintenance.
- **Scalability:** It offers scalability, enabling easy adjustments to accommodate organizational growth or changing requirements over time.
- **Customization:** The environment can be customized to align with Hawaii's unique workflows, processes, and requirements, ensuring optimal integration and functionality.

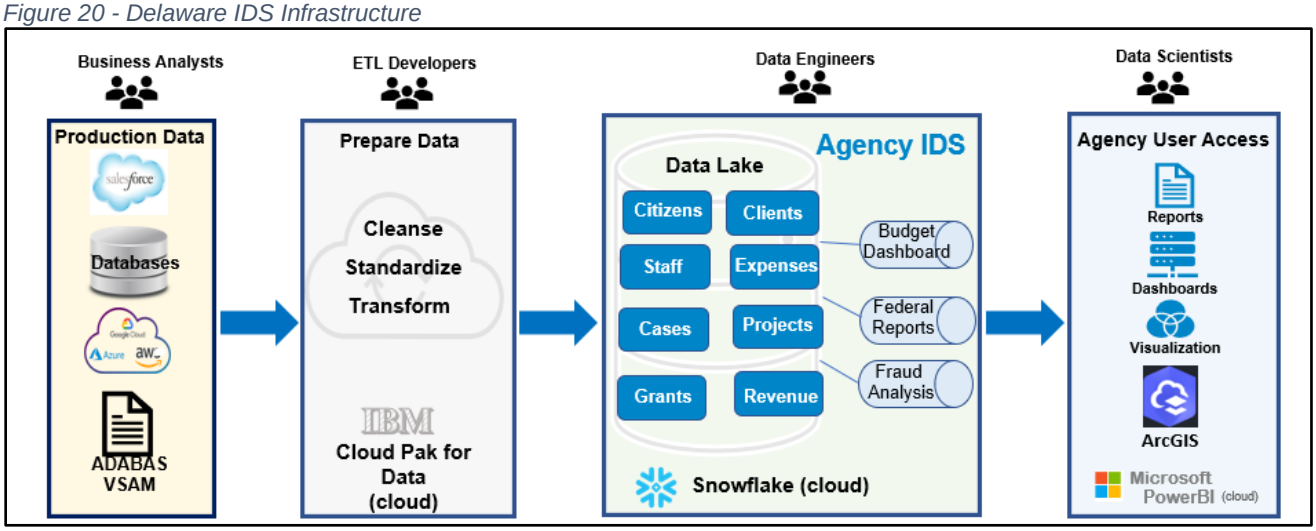
Figure 19 - HAP Data Environment



Delaware Integrated Data System

Error! Reference source not found. below shows the foundational infrastructure that has been successfully implemented to facilitate secure and cost-effective data sharing both within and across various agencies, once DIDS id fully functional. This infrastructure operates by processing production data into the Integrated Data Store (IDS), from which different use cases can access the prepped data. To ensure privacy and compliance, all personally identifiable information (PII) data is masked within Snowflake and remains masked wherever it is utilized. Additionally, production data is regularly refreshed within the data lake, ensuring that the information available is always up-to-date and reliable for analysis and decision-making purposes.

Figure 20 - Delaware IDS Infrastructure



CHAPTER 4: FUNDING AND SUSTAINABILITY

Introduction

Integrated Data Systems should consider all available and prospective revenue sources to support their initial development and implementation, and then ongoing operations thereafter. While the VT HWDC has relied solely on a state appropriation to embark on the implementation of the HWDC,, this funding source may not be sufficient to cover all program costs in perpetuity. Developing a sustainable and diversified funding plan at the outset of the HWDC will provide stability to the project team and promote stakeholder buy-in and engagement to a longstanding program.

Key Takeaways:

- Explore available and prospective revenue sources to support the development of the IDS and ensure program sustainability.
- Potential data sources include:
  - State appropriations
  - Revenue-generating data release products
  - Assessments/ongoing subscriptions for data access
  - Federal funds
  - Private grants, trusts, and foundations
- Start early as the processes involved in accessing funds may take longer than expected.

Potential Funding Sources

State Appropriations

Most IDS programs receive core funding from a state appropriation to support operational costs, in full or in part. This state appropriation is typically provided through a state's General Revenue budget and follows the state's standard budgeting cycle – including preparation by the State's Budget and Management Division, approval by the Legislature, and passage of the associated appropriations bill.

Programs requesting state funding must be aware of the state's budgeting cycle which may require proposals to be submitted a year in advance of the start of the budget cycle. In VT, for a program to receive funding for the state fiscal year (which begins on July 1), the proposal needs to be included in the funding request that is initiated in July of the previous year (i.e., the SFY26 budget (July 1, 2025 – June 30, 2026), is initiated by requests submitted in July 2024).

Revenue-Generating Data Release Products

In addition to a state appropriation, many IDS programs have a data request process through which data that is not otherwise publicly available can be requested for a fee. IDS programs determine pricing for these revenue-generating products based on what is allowable according to their enabling laws, regulation, or internal policy. Considerations include the cost of producing and transmitting the data, maintaining the database, providing user support, and other states or programs fees for similar data. Pricing for data products also depends on the data product requested and usually, the type or "category" of user requesting it. Some states use a cost

template to provide a customized quote for each request, whereas others have fixed fees for all requests of the same type. Some states also charge an application fee to cover the costs of reviewing applications to deter frivolous applications.

In determining whether revenue-generating data release is appropriate for the VT HWDC, the state should consider who the intended users of the database will be and whether developing a fee schedule aligns with the mission and vision for the IDS.

If data is available to non-state agency requestors for a fee, similar programs have sought authority to establish a restricted receipt account that can accept revenue from non-governmental sources and restrict use of such funds to the IDS program. New legislation may be needed for Vermont to be eligible for this option.

Assessments on Data Users

Assessments on primary data users are similar to revenue-generating data products, but describe a regular assessment (e.g., annual subscription fee) on certain ongoing data users such as hospitals, educational institutions, workforce training programs, or other entities that will steadily access the database. Like ad-hoc revenue generating data products, assessments cover the costs of maintaining and operating the IDS resource.

Federal Funding Opportunities

The federal government provides various funding opportunities that can support IDS initiatives. These funding opportunities vary and include grant-making programs that address specific federal priorities (e.g., CDC Public Health Infrastructure Grants to recruit, retain, and train healthcare workers) or public health events (e.g., American Rescue Plan grants to state health departments to address healthcare access issues exacerbated by the COVID-19 pandemic). federal funding opportunities can be found in the *Compendium of Federal Data Sources to Support Health Workforce Analysis* prepared by the Health Resources and Services Administration (HRSA) of the US Bureau of Health Workforce.[3]

One of the most viable sources of sustainable funding to support an IDS comes in the form of Federal Financial Participation (FFP), also known as “Medicaid Match.” Medicaid Match generally comes in two types: administrative match[4] and enhanced match[5]. All states that operate their Medicaid programs within federal guidelines are entitled to federal reimbursement for a share of their total program costs. The federal share for Medicaid administrative activities – known as the administrative match - is generally 50%. This 50% match covers anything found “necessary for the proper and efficient administration of the State Medicaid plan”. In some instances, however, CMS provides a higher matching rate for select, federal priority services or

[3] https://bhw.hrsa.gov/sites/default/files/bureau-health-workforce/data-research/compendium.pdf
[4] Soc. Sec. Act Sec. 1903(a)(7). Found at: https://www.ssa.gov/OP_Home/ssact/title19/1903.htm
[5] Soc. Sec. Act Sec. 1903(a)(3). Found at: https://www.ssa.gov/OP_Home/ssact/title19/1903.htm

projects, including upgrades to Medicaid computer and claims processing systems[6], projects which support Medicaid data integration and data sharing[7], and data resources that can assist in identifying Social Determinants of Health[8]. This higher match rate for specific programs or activities is known as the enhanced match rate. Under the enhanced match rate, activities/programs can be matched at 90% while they are being developed, and 75% to maintain and operate. In other words, for every dollar expended during development and implementation, CMS will provide 90 cents and the state must provide the remaining 10 cents from nonfederal funds (usually state appropriations).

The amount of FFP received for an IDS initiative varies. Some states have received FFP just for the development of Medicaid dashboards or reports out of their IDS, while other states receive FFP for all program costs, as even non-Medicaid data is needed by the Medicaid program for benchmarking purposes. FFP must be requested by the state's single Medicaid agency and is approved on a rolling-basis for two federal fiscal years at a time – with the opportunity to submit annual project and budget request updates.

Partnering with the state's Medicaid department on an Advanced Planning Document (APD) request to receive FFP is an important route for all IDS teams to consider, as it can serve as a continuous source of substantial funding. Below are a few tips to consider before embarking on a FFP request for an IDS:

1. **Review Recent CMS Guidance:** Recent guidance has paved the way for integrated databases applying for enhanced federal financial participation. States should review 42 CFR Part 433, the CMS State Medicaid Directors Letters from 2016[6], 2018[7], and 2021[8] and Jessica Kahn's presentation to the National Committee on Vital and Health Statistics[9].
2. **Develop Strong Partnerships**: Build close working relationships with members of the state's Medicaid team and the regional Centers for Medicare and Medicaid (CMS) office.
3. **Start Early**: Cultivating relationships, drafting the APD proposal and awaiting CMS approval can each take longer than expected. Be prepared for delays and take advantage of the relationships and learnings gained along the way.
4. **Make Clear Connections to Value:** Develop strong use cases for how the IDS data would benefit Medicaid operations, analyses, and reporting. As always, leave the jargon at home and focus on the what, why and how.
5. **Create a Shared Narrative**: It is crucial to make every partner organization and department have a shared understanding of the IDS' opportunities, priorities, and guardrails. Shaping these key messages into common talking points will demonstrate the cohesiveness of your plan and provide the best opportunity for effective communications with CMS.

[6] https://www.medicaid.gov/federal-policy-guidance/downloads/smd16010.pdf
[7] https://www.medicaid.gov/federal-policy-guidance/downloads/smd18005.pdf
[8] https://www.medicaid.gov/federal-policy-guidance/downloads/sho21001.pdf
[9] http://www.ncvhs.hhs.gov/wp-content/uploads/2016/05/Panel-2-Jessica-Kahn-CMS-Written-20160June17.pdf

6. **Prepare for Questions**: CMS will have a lot of specific questions for you to answer, so it is important to be prepared for the details required to answer them. Consider developing a list of possible questions you might receive and think about ways you might answer them. Share your thoughts with your partners and see if they would answer them similarly.

Private Grants, Trusts, and Foundations

In addition to securing a state appropriation and exploring user fees and federal funding sources, several IDS programs have successfully procured additional, private grant dollars to offset the remainder of their operational costs. There are countless private grants, trusts, and foundations that can serve as a source of funding. Grants usually fund and support specific projects, analyses, or reports (e.g., reports specific to certain diseases, SDOH factors, etc.) and not infrastructure. For this reason, grants are not recommended as a viable funding source during IDS development, but they are able to support report production in a mature system.

Examples of private grants include the Robert Wood Johnson Foundation which typically supports analyses that highlight health equity and healthcare access issues and college grants/endowments that focus on current state priorities and providing a data source for graduate students to conduct research.

Appendix A

Best Practices – Developing an Integrated Data System – Weekly Wednesday Meetings
1/10/24 – 2/28/2024
4:00-5:00 PM

Meeting #	Meeting Date	Proposed Topic	Speaker
1	1/10	**Overview of IDS' nationwide, including state-sponsored Healthcare Workforce data initiatives.** Including: years of operation, implementation timeline, host agency, contributing data sources, funding mechanism, allowable use cases, and stakeholders. To include: 1. Rhode Island Ecosystem 2. Hawaii Health Analytics Platform 3. Delaware Integrated Data System 4. Washington Workforce Board 5. California Dept. of Health Care Access and Information's (HCAI's) Health Workforce Research Data Center 6. Virginia Dept. of Health Professions' Healthcare Workforce Data Center 7. Texas Health Professions Resource Center	Tanya Bernstein, Executive VP, FHC
2	1/17	**Best Practices in Getting Buy-In from Contributing Data Sources ("Data In")** • Purpose/Allowable Use Cases • Data Sharing Agreements • Data Governance Roles	Saral Linberg, Senior Data Science Consultant, FHC Sely-Ann Johnson, Project Manager, FHC
3	1/24		Rebecca Lebeau, Director of Analytics, RI Data Ecosystem
4	1/31	**Best Practices in Responsible Data Access ("Data Out")** • Available Data and Data Products • **Sensitive/Masked Information** • Data Request Process • Data Use Licenses3we • Role-Based Access Control (RBAC)	Tanya Bernstein, Executive VP, FHC Sely-Ann Johnson, Project Manager, FHC
5	2/7		
6	2/14	**Best Practices in Data Architecture** • How to promote interoperability • How to promote modularity • Keeping scalability in mind • On prem vs. cloud considerations • Firewalls/partitions to protect PII and PHI	Saral Linberg, Senior Data Science Consultant, FHC Gabriella Caldas, Data Engineer, FHC

7	2/21	**Staffing Considerations** • Necessary roles • Necessary skills • In-house vs. contracted staff • Estimated levels of effort by year	Ranjani Starr, Direct or Analytics, HI Medicaid
8	2/28	**Funding Models** • How other IDS' stay afloat. • Fee structure • General Revenue approach • Medicaid Match	Tanya Bernstein, Executive VP, FHC