

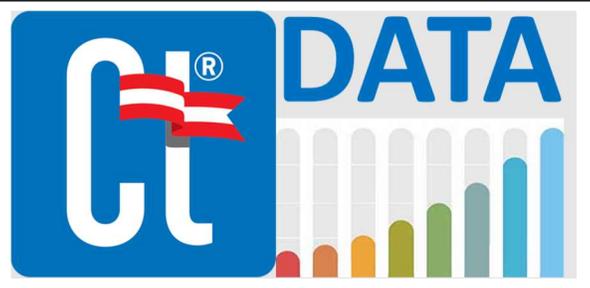
Connecting the people and processes involved with data to promote communication between, and integration of, formerly siloed data, teams, and systems within executive branch agencies.

# Draft State Data Plan

Submitted pursuant to Public  
C.G.S. Sec. 4-67p

Office of Policy and Management  
November 9, 2020

---



# Table of Contents

---

- Introduction ..... 2
- Mission..... 2
- Scope..... 2
- Roles and Responsibilities..... 3
- About the Process..... 3
- Progress and Lessons Learned ..... 4
- Principles..... 4
- Focal Points..... 6
- Goals and metrics ..... 11
- Glossary of Terms ..... 12

## Introduction

---

The Connecticut State Data Plan serves as a framework for the state's executive branch agencies to engage in a consistent approach to data stewardship, use, and access. The State Data Plan *is not just an open data plan* but is applicable to all data in the custody and control of executive branch agencies. The plan was required to be developed in accordance with [C.G.S. § 4-67p](#). Specifically, the state data plan shall:

- establish management and data analysis standards across **executive branch agencies**;
- include specific, achievable goals within the two years following the adoption of such plan, as well as longer term goals;
- make recommendations to enhance standardization and integration of data systems and data management practices across **executive branch agencies**;
- provide a timeline for the review of any state or federal legal concerns or other obstacles to the internal sharing of data among agencies, including security and privacy concerns; and
- set goals for improving the open data repository.

The plan is required to be updated every two years.

The statutes also require that *information technology-related actions and initiatives of executive branch agencies must be consistent with the plan*. Further, the Chief Data Officer is required to establish procedures for each agency to report on its progress toward achieving the goals articulated in the plan. A supplement to the plan will report progress during 2019 – 2020.

The inaugural plan was organized around **Principles** which represent a framework under which state agencies should organize and operate; **Focal Points** which represent areas where agencies should emphasize the sharing, integration, and availability of data; and **Goals** which represent the desired outcomes of plan implementation.

The 2021 – 2022 plan uses data gathered during the inaugural plan to narrow and refine the **Focal Points** leading to a revised set of **Goals**.

## Mission

The purpose of this plan is to connect the people and processes involved with data to promote communication between, and appropriate integration of, formerly siloed data, teams, and systems. It seeks to promote process change, organizational realignment, value congruence, and the appropriate use of technology to facilitate relationships between everyone who handles and uses data.

By aligning the data related work of agencies with the plan's Principles, agencies should begin to realize efficiencies and process improvements that will ultimately reduce the amount of overhead necessary to leverage their data assets. Agencies must begin to closely connect the people who provide, collect and prepare the data, those who analyze the data, and those who put the findings from those analyses to good use.

## Scope

The plan to supports agency efforts and to brings those efforts into alignment to provide increased efficiency in the use of limited resources for data stewardship, use, and access. The plan goals and

implementation steps provide a framework to support the use of data to improve agency outcomes and performance. The plan does not require agencies to dedicate additional resources to its implementation, nor does it generally require agencies to begin collecting or creating new data.

While state agencies collect and manage data to operate programs and services and have used this information for reporting for decades, only recently have state officials begun to harness administrative records for analytical purposes to achieve desired goals or outcomes. Although this plan does not specifically address variations in the sensitivity of interagency data, it broadly addresses many underlying challenges associated with data collection and provides a standard framework that can be applicable across executive branch agencies.

## Roles and Responsibilities

In addition to establishing the State Data Plan, [C.G.S. § 4-67p](#) establishes the position of Chief Data Officer within the Office of Policy and Management, and requires each executive branch agency to designate an agency data officer, each with roles and responsibilities outlined below. Further, the act requires that each agency conduct an inventory of its “high-value data” annually and formulate plans to enhance the availability of open data, known as “open data access plans.” The following are the roles and responsibilities of the Chief Data Officer and Agency Data Officer that are established by [C.G.S. § 4-67p](#):

### Chief Data Officer:

- Directing executive branch agencies on the use and management of data to enhance the efficiency and effectiveness of state programs and policies;
- Facilitating the sharing and use of executive branch agency data (A) between executive branch agencies, and (B) with the public;
- Coordinating data analytics and transparency master planning for executive branch agencies;
- Creating the state data plan; and
- Providing a procedure for each agency head to report regarding the agency's progress toward achieving the plan's goals.

### Agency Data Officer:

- Coordinating and submitting agency high value data inventories annually;
- Coordinating and submitting agency open data access plans;
- Serving as the main contact person for inquiries, requests or concerns regarding access to the data of such agency; and
- Establishing procedures to ensure that requests for data that the agency receives are complied with in an appropriate and prompt manner in consultation with the Chief Data Officer.

It should be noted that, while [C.G.S. § 4-67p](#) requires Agency Data Officers to establish procedures to ensure that requests for data that the agency receives are complied with in an appropriate and prompt manner, as well as establish open data access plans, the law does not create a separate entitlement or an alternative to the Freedom of Information Act process.

## About the Process

The 2021 – 2022 State Data Plan will be developed through an iterative process, with opportunities for both public and agency input as follows:

**Preliminary draft** – Present a draft of the State Data Plan to the Data Analysis Technology Advisory (DATA) board pursuant to [C.G.S. § 4-67p](#) and [C.G.S. § 2-79e](#). The DATA board shall hold a public hearing on the plan and submit any recommendations within 30 days of receiving the draft plan.

**Final Plan** – Following the public hearing and after receipt of public comment via the [State Data Plan website](#), a final revised plan will be issued by December 31, 2020.

## Progress and Lessons Learned

---



The conclusion of the first State Data Plan presents an opportunity to reflect on lessons learned over the past two years. A supplemental document will report examples of progress related to the Principles and Focal Points from the inaugural [State Data Plan](#).

## Principles

---



The twelve Principles below represent an ongoing framework for agencies to mature and enhance their management, use, sharing, and analysis of data. The Principles remain the same as in the inaugural 2019 – 2020 plan as they should serve as guidelines for continuous improvement and to inform decision-making on an ongoing basis. The current plan shifts the agency practices and areas for CDO support to the Goals of the plan. A supplement to the plan will describe successes and lessons learned over the last two years for the Principles, agency practices and CDO support.

### Data Governance and Quality

- ① Appreciate that the data we collect require a considerable investment of resources and have value beyond the purpose for which they are collected.
- ② Perform continuous data quality and analytics improvement to ensure the value of data is protected and maximized.
- ③ Ensure standardized data governance in order to protect data and improve its quality and utility.
- ④ Create, acquire, use, and disseminate data deliberately and thoughtfully; in compliance with federal law and state statute, and considering, quality, consistency, privacy, equity, client data dignity, value, reuse, and interoperability from the start.
- ⑤ Coordinate and prioritize data needs and uses, utilize data from multiple sources, and acquire new data only when necessary.

### Privacy and Ethics

- ⑥ Protect individual privacy and maintain confidentiality using effective data stewardship and governance, and by maintaining modern data security practices and technology.
- ⑦ Employ ethical standards in the use, analysis, sharing, and integration of data to avoid intrusion into the lives of Connecticut residents, and disparate impact.

### Data Sharing Between Agencies

---

- ⑧ Improve data sharing and access with ongoing input from users and other stakeholders, including those whose personal and protected data are collected in state agency systems.
- ⑨ Create clear and predictable pathways for data sharing which are necessary for effective data use and sharing.
- ⑩ Manage a data asset one time and use it for multiple purposes, to the extent possible, within legal and regulatory constraints.

### Culture of Continuous Learning and Collaboration

- ⑪ Promote a culture of continuous and collaborative learning, with data and about data.
- ⑫ Embrace openness, transparency, and accountability.

## Focal Points

---



The Focal Points represent areas where state agencies will emphasize the sharing, integration, and availability of data. The Focal Points represent important topics where data from multiple state agencies are necessary to drive effective policy, or to inform the effective and efficient delivery of programs and services. Given that this plan is limited to specific executive branch agencies, areas that would require data from the judicial branch or Constitutional offices have not been included. All efforts undertaken within these Focal Points should be carried out in accordance with the Principles as articulated by the plan.

Three related areas will be focal points for the 2021 – 2022 State Data Plan:

- 1) COVID-19 response and recovery
- 2) Equity in the data lifecycle
- 3) Using data to inform decision-making

Future drafts of the plan will provide supporting data for the three focal points. Goals and implementation steps will be based on this data.

---

### COVID-19 response and recovery

---

The COVID-19 pandemic has brought into focus the need for state agencies to provide timely, accurate data to the public and to quickly and effectively use data to inform policy and decision-making. Agencies have developed new data collections, new systems, new interagency agreements, new analysis and new tools in response to the pandemic. State agencies, led by the Department of Public Health, have refined an automated process to post daily indicators and metrics on cases, testing, hospitalizations and deaths. While it is uncertain how long the COVID-19 pandemic will impact daily life, the crisis has prompted permanent, structural changes across systems and data management.

#### ***Supporting data:***

Since the start of the COVID-19 pandemic, state agencies, led by the Department of Public Health, have worked to develop, automate and refine a process for daily updates on a wide range of indicators and metrics. Traffic to the related assets on the open data portal has increased by roughly four-fold. The COVID-19 data story has received over 800,000 direct views since the start of the pandemic, and a data story featuring schools data received 100,000 within the first week after launch, roughly equivalent to a month's traffic to the entire open data portal pre-COVID.<sup>1</sup>

Public engagement has similarly increased - the portal has received more data requests since March 2020 than were posted in the first six years the portal was online, from an average of one suggestion every two weeks to one per day.<sup>2</sup> COVID-19 datasets have increased API calls to the portal similarly, averaging around 500,000 / month since March, roughly twice the number before March 2020.

---

<sup>1</sup> Analysis of site analytics for data.ct.gov

<sup>2</sup> Analysis of dataset suggestions from: <https://data.ct.gov/nominate/>; 207 suggestions were posted from March 26, 2020 to November 6, 2020, and 143 were posted between March 2014 and March 2020.

---

Connecticut’s open data efforts have been able to support the increased demands for agency data but have also identified areas for improvement for 2021 – 2022. Increased usage of open data assets has not been universal. Roughly 70 percent of the datasets published on the portal have not been updated in 2020, and just over 30 percent have not been updated since 2015.<sup>3</sup> Only around 17 percent have been updated within the last 30 days, implying they are updated at least monthly.<sup>4</sup> Most datasets have a limited number of views or downloads – with notable exceptions for several COVID-19 datasets and data related to consumer protection - although traffic is not a direct measure of impact.<sup>5</sup> Roughly one quarter of the datasets on the portal do not have agency attribution, limiting the ability to measure publication and impact by agency.<sup>6</sup>

The pandemic response effort has identified other areas for improvement in agency systems and the ability to link and analyze data across agencies.<sup>7</sup> The Lamont administration has said that “We’ve had to piece together public data, private data, data from the Connecticut Hospital Association, and other information to understand what’s going on...Tools like [a health information exchange would] prevent that time from being wasted. It would have been easier ... to have something like this in place seven months ago.”<sup>8</sup>

***Potential implementation steps:***

The following are potential actions for executive branch agencies for this focal point:

- Develop an approach to prioritize high-value datasets for the open data portal, and policies and procedures to improve or, when necessary, remove or deprecate datasets that are out-of-date or that adversely impact usability of the portal
- Improve the user experience with better navigation, user-friendly tools (like data stories), more consistent documentation and metadata and schedules for the release of high-value datasets
- Improve the ability to provide daily or real-time updates through automated feeds to the portal
- Support the development of health data infrastructure for continued response to the COVID-19 pandemic, and for future emergencies
- Coordinate fully with planning efforts for health information technology and state information technology infrastructure
- Improve access to, coordination and management of the foundational geospatial data that is critical for response to future emergencies

---

## Equity in the data lifecycle

---

The COVID-19 pandemic has demonstrated, again, the disparities in outcomes experienced along many dimensions – particularly by race and ethnicity, but also by gender, age, ability status and other factors. An

---

<sup>3</sup> Analysis of public datasets on data.ct.gov, as of November 5, 2020: <https://data.ct.gov/dataset/data-ct-gov-public-assets/k9rk-977m>

<sup>4</sup> Analysis of public datasets on data.ct.gov, as of November 5, 2020: <https://data.ct.gov/dataset/data-ct-gov-public-assets/k9rk-977m>

<sup>5</sup> Analysis of public datasets on data.ct.gov, as of November 5, 2020: <https://data.ct.gov/dataset/data-ct-gov-public-assets/k9rk-977m>

<sup>6</sup> “CT Open Data: Measuring Open Data Impact and Use.” <https://portal.ct.gov/-/media/CT-Data/Open-Data-Impact.pdf>

<sup>7</sup> “As it prepares to reopen, Connecticut’s count of COVID-19 deaths is fragmented and incomplete.” CT Mirror, May 19, 2020. <https://ctmirror.org/2020/05/19/as-it-prepares-to-reopen-connecticuts-count-of-covid-19-deaths-is-fragmented-and-incomplete/>; “CT catches up on delayed federal unemployment relief.” CT Mirror, September 29, 2020. <https://ctmirror.org/2020/09/29/ct-catches-up-on-delayed-federal-unemployment-relief/>

<sup>8</sup> “As pandemic wears on, Connecticut prepares to launch its long-awaited health information exchange,” CT Mirror, October 15, 2020. <https://ctmirror.org/2020/10/15/as-coronavirus-lingers-ct-prepares-to-launch-its-long-awaited-health-information-exchange/>

---

intentional focus on the data lifecycle – from planning, through collection, management, analysis and dissemination – can improve the ability to achieve equitable outcomes. All aspects of the data lifecycle merit attention, across agencies, as disparities in housing, education, criminal justice remain.

***Supporting data:***

It is difficult to directly measure whether the use of data promote equity, particularly racial equity. However, population data for Connecticut shows us areas in which we can improve how data are used to support equitable outcomes.

COVID-19 age-adjusted case and fatality rates are two-to-three times higher for Black and Latino residents of Connecticut, than for White residents.<sup>9</sup> About one-third of cases are still missing data on race / ethnicity, although data are reported for the race / ethnicity, gender, age and other dimensions.

Survey data shows that “Latino residents reported being laid off at higher rates than whites or Blacks, with 37% of residents saying that they had lost their jobs. Latino residents also reported thrice the rate of housing insecurity when compared to white people.”<sup>10</sup> Connecticut has two of the top 20 most segregated metro regions in the country (Bridgeport and Hartford).<sup>11</sup>

***Potential implementation steps:***

The following are potential actions for executive branch agencies for this focal point:

- Pursue opportunities for training and capacity-building for state agency staff
- Improve standards for data collection and dissemination to ensure that categories for gender identity, race / ethnicity and other factors are up-to-date and consistent across agencies
- Encourage consideration of racial equity and disparate impact in the process used to access, request and publish data
- Increase reporting and disclosure of data by factors such as race, ethnicity and gender, and their intersections, within the appropriate guidelines to preserve privacy
- Improve the tools for reporting and disseminating data to reach a wider and more diverse set of audiences
- Increase the depth and breadth of avenues for resident involvement in the data lifecycle
- Improve access to, coordination and management of the foundational geospatial data that is critical for understanding spatial disparities

---

## Using data to inform decision-making

Whether it is called being data-driven, [data-informed](#), [results-driven](#), [evidence-based](#), [performance-informed](#), results-focused or any other compound adjective, state administrative data can be more actively used to inform decision-making, including for policy development and the delivery of services. This ambitious focal point requires progress on several components:

---

<sup>9</sup> Analysis of case and deaths data from data.ct.gov: <https://data.ct.gov/Health-and-Human-Services/COVID-19-rates-by-race-ethnicity/qd9u-qr98>

<sup>10</sup> Analysis of DataHaven Wellbeing Survey results in “State of Wellbeing’ survey quantifies racial inequities in both health and economy.” CT Mirror, September 17, 2020. <https://ctmirror.org/2020/09/17/state-of-wellbeing-survey-quantifies-racial-inequities-in-both-health-and-economy/>

<sup>11</sup> “Residential Segregation Data for U.S. Metro Areas.” Governing, March 24, 2019. <https://www.governing.com/gov-data/residential-racial-segregation-metro-areas.html>

- 
- ***Integrated data system:*** A core building block to use data to inform decision-making is the development of a system to integrate data across state agencies. As described in a report on the legal issues in interagency data sharing:

“Today, individuals who receive governmental services are often involved with multiple systems. For example, a young mother and her child may receive SNAP benefits, childcare subsidies, child support payments, or other state-delivered services. Each of these programs was designed to fulfill a distinct purpose, and each collects different data and follows different rules and requirements. Each program’s database only identifies patterns or characteristics of those served within that particular agency or program. Isolated databases omit information from other agencies or programs that could be analyzed to increase wellbeing, long-term personal success and reduce costs to state and local government.”<sup>12</sup>
  - ***Enterprise legal and governance framework:*** The foundation for an integrated data system is the legal and governance framework. Legal agreements are not barriers or obstacles, they are supports to the safe, ethical and secure sharing of data across agencies boundaries. The absence of statewide governance to coordinate data sharing efforts leads to fragmented approaches to sharing data on high-priority issues which reduce the ability of the state to mobilize a response.
  - ***Increased analytical capacity:*** Turning data into information and insight requires people. In the current fiscal environment, state agencies need to work collaboratively to make efficient use of internal resources and to effectively partner with outside analytical capacity.
  - ***Whole-person perspective on data:*** Using data to improve delivery of services has great potential to improve the lives of residents, but can only occur through the safe, ethical and secure linking of information across agencies. As part of the 2 Generational Initiative Interagency Plan,<sup>13</sup> OPM identified “two approaches to coordinating existing resources – identifying established programs and identifying and monitoring proofs of concept.”<sup>14</sup> At the same time, a ‘whole-person’ data view cannot be developed without attention to the ethical and equity issues inherent if data are used beyond the purpose for which they were originally collected.

***Supporting data:***

Given fiscal constraints and an impending ‘retirement cliff,’<sup>15</sup> the state must find ways to make efficient use of limited resources and has identified the development of improved digital services and ‘one-stop’ solutions as part of a focus on increased efficiency. State leaders have said that with “better information sharing, agencies not only could avoid duplicative efforts, but also coordinate and improve services.”<sup>16</sup> A data matching effort for 10,000 individuals and families experiencing homelessness found that, on average, they were involved with four of the six agencies that participated in the data match.<sup>17</sup>

---

<sup>12</sup> [Legal Issues in Interagency Data Sharing, Report for Public Act 19-153](#), Office of Policy and Management, January 2020

<sup>13</sup> [2 Generational Initiative Interagency Plan](#), Office of Policy and Management, December 2019

<sup>14</sup> [2 Generational Initiative Interagency Plan](#), Office of Policy and Management, December 2019

<sup>15</sup> “Boston Firm That Mapped CT’s COVID Reopening Will Craft Major Downsizing of State Agencies.” The CT Mirror, September 25, 2020. <https://ctmirror.org/2020/09/25/boston-firm-that-mapped-cts-covid-reopening-will-craft-major-downsizing-of-state-agencies/>.

<sup>16</sup> “Lamont Moving Forward with Agency Overhauls as Wave of Retirements Approaches,” The CT Mirror, September 2, 2020, <https://ctmirror.org/2020/09/02/lamont-moving-forward-with-agency-overhauls-as-wave-of-retirements-approaches/>.

<sup>17</sup> Preliminary analysis by Beacon Health Options for [Governor’s Task Force on Housing and Supports for Vulnerable Populations](#), September 2020.

---

Interviews with state agency staff for the Data Sharing Playbook<sup>18</sup> identified areas for process improvement and clarity on how data can be shared and used. Agency staff often feel they need to “start from scratch with each data sharing project,” with challenges identified in partnership and staffing, the legal framework and data standards across agencies. Sentiments from interviews included<sup>19</sup>:

- “Stumbling blocks are the requester knowing what they want, if they know what data is available, either in a data dictionary or record layouts.”
- “The strategy needs to include a legal framework – a lot of lawyers in agencies are not willing to get into data sharing of protected or what they view as protected.”
- “[Other agencies may say] I’m worried that if you use my data and come out with statements on our program from a different view, that’s a loss of control I’m not comfortable with”

Despite the perceived challenges, data sharing does occur. The development of actionable intelligence from administrative data begins with the legal foundation.<sup>20</sup> A proliferation of data sharing agreements makes oversight difficult and reduces the ability to protect clients’ data and manage risk. A survey of state agencies found 224 separate data sharing agreements across 17 agencies. Ninety percent of the agreements were two-party agreements, primarily for sharing individual-level data. Only three percent were agreements of more than three parties. The same agencies identified a thicket of 136 different laws and regulations that governed agency data sharing, with three agencies identifying more than 25 different laws and regulations for their agency. Eighty percent of the laws and regulations listed in the survey were at the state level (109), while 20 percent of the laws and regulations listed were at the federal level (27).<sup>21</sup>

One state effort to integrate data is the P20 WIN (Preschool through 20 Workforce Information Network) system. The results of P20 WIN data matches have been cited nationally and used within the state.<sup>22</sup> P20 WIN integrates data from education and workforce agencies, based upon data requests from state agencies and external partners. Since 2013, P20 WIN has processed 20 requests, an average of just under three requests per year.<sup>23</sup> Six requests have been from external researchers (ex. Urban Institute, Yale University), and fourteen have been ‘internal’ requests between state agencies.<sup>24</sup> The turnaround time for a request has been more than 8 months on average, from the time a data request is finalized to the pre-publication review of results.<sup>25</sup>

---

<sup>18</sup> [CT Data Sharing Playbook](#), “[Want to improve data sharing? Start with a playbook](#),” GovBloom / Skylight, Medium.com, April 2020

<sup>19</sup> Data Sharing at Connecticut State Agencies, Team Skylight preliminary research findings, December 2019

<sup>20</sup> Fantuzzo, John, Culhane, Dennis P. (Eds.) 2015. “Actionable Intelligence: Using Integrated Data Systems to Achieve a More Effective, Efficient and Ethical Government.” New York: Palgrave Macmillan US. Accessed at: [https://www.aisp.upenn.edu/wp-content/uploads/2015/03/Chapter1\\_ResourcesPg.pdf](https://www.aisp.upenn.edu/wp-content/uploads/2015/03/Chapter1_ResourcesPg.pdf)

<sup>21</sup> [Legal Issues in Interagency Data Sharing, Report for Public Act 19-153](#), Office of Policy and Management, January 2020

<sup>22</sup> Description and resources found at: <https://portal.ct.gov/OPM/P20Win/>

<sup>23</sup> [P20 WIN reports](#), accessed November 2020. Dates are based on the signatures for data request forms and pre-publication review. Includes only those requests that had reached pre-publication review by the time of this plan; two additional requests are in process at the time of the plan.

<sup>24</sup> Analysis of [P20 WIN data request documents](#). ‘Internal’ requests were those from members of the P20 WIN Data Governing groups; external requests were from outside participating agencies or faculty from participating agencies that do not directly participate in P20 WIN request review and approval.

<sup>25</sup> Analysis of [P20 WIN data request documents](#). The time was measured from the last signature on the data request form to the last signature date on the pre-publication review form. Requests may take additional time to develop a formal request that is available for signature. Some data requests are repeat or annual requests, so this understates the actual number of data matches completed.

---

Staff availability and capacity is a binding constraint on managing interagency data requests, particularly ‘internal’ requests for cross-agency analysis. As of November 2020, state agencies have roughly 60 research analysts on staff, in total, and only four research supervisors, limiting the career pathways in government for analysts.<sup>26</sup>

**Potential implementation steps:**

The following are potential actions for executive branch agencies for this focal point:

- **Integrated data system and enterprise legal and governance framework**
  - Continue work to implement the uniform interagency data sharing protocol, developed as part of state 2Gen efforts
  - Increase efficiency through process improvements, such as development of the Data Sharing Playbook
  - Increase transparency of agency-level efforts at performance management and work towards common standards for implementation
  - Build on cross-agency efforts at performance management, particularly from the Governor’s Workforce Council, to support centralized data on program performance and rigorous evaluation of programs, where possible
- **Increased analytical capacity**
  - Identify opportunities for skill-building and career pathways for existing analytical staff
  - Improve the ability to engage in external research partnerships, through the development of research agendas and priorities, and supporting materials, like data dictionaries, legal guidelines and repositories of past work
  - Consider pursuing external partnerships particularly to support ‘mixed methods’ approaches, mixing qualitative and quantitative data
- **Whole-person perspective on data**
  - Continue to expand the P20 WIN data system to cover new agencies and datasets, including social services, homelessness, financial aid and child welfare
  - Continue efforts to support and expand the secure master person index infrastructure, in collaboration with [Connie](#), the health information exchange, and health and human services agency partners, creating the necessary infrastructure for data integration
  - Align state efforts to improve delivery of services, optimize information technology resources and to expand the P20 WIN system

---

## Goals and metrics



The overarching purpose of the State Data Plan is to facilitate the lawful and responsible sharing and use of data amongst executive branch agencies in order to enhance the efficiency and effectiveness of state policies and programs. This must occur in a manner that recognizes and supports the rights, privacy, and protection of clients/consumers whose data are collected in state data systems. The

following goals, when pursued collectively amongst agencies, will advance this objective.

---

### ① Improve and increase relevance of the open data portal and related tools

Progress on this goal could be measured through metrics such as:

---

<sup>26</sup> Analysis of [State Employee Payroll Data Calendar Year 2015 through Present](#), for the following job classifications: Research Analyst, Associate Research Analyst, Research Analyst Supervisor.

- 
- Site traffic, including to data stories and dashboards and other tools to reach a wider audience
  - Frequency and consistency of updates, particularly to high-value or high-priority data
  - Specific instances of use of agency data, particularly open data, to inform policy and practice

## ② Ensure that the data lifecycle promotes equity, particularly racial equity, wherever possible

We cannot move the needle on population stats solely by improving agency data management. However, we can track progress on this goal through metrics such as:

- Agency participation in training or capacity-building efforts
- Increased consistency in collection and reporting of demographic data across agencies
- Increased number and percentage of datasets that are disaggregated by factors like race, ethnicity, gender, etc., and by the intersection of one or more of those factors
- Improved availability of metadata and documentation for high-value datasets
- Specific instances of change in state or agency practice (ex. development of resident voice)

## ③ Improve use of data to inform decision-making

Progress on this goal could be measured through metrics such as:

- Reduced time to complete interagency agreements, improved customer experience for data requestors, increased number of external research partnerships
- Tracking agency participation in training or capacity-building efforts
- Ability to identify performance metrics and evidence base for critical programs

---

## Glossary of Terms

---

**Data:** The final version of statistical or factual information that: (A) is reflected in a list, table, graph, chart or other non-narrative form that can be digitally or nondigitally transmitted or processed; (B) is regularly created or maintained by, or on behalf of, an executive branch agency; and (C) records a measurement, transaction or determination related to the mission of the agency or is provided to the agency by third parties pursuant to law.

**Data dignity:** The data subject's human dignity, legitimate interests and fundamental rights, with particular regard to the transparency of processing, or the transfer of personal data.

**Executive branch agency:** Any agency listed in section 4-5 of the general statutes.

**High value data:** Any data that the department head determines (A) is critical to the operation of an executive branch agency; (B) can increase executive branch agency accountability and responsiveness; (C) can improve public knowledge of the executive branch agency and its operations; (D) can further the core mission of the executive branch agency; (E) can create economic opportunity; (F) is frequently requested by the public; (G) responds to a need and demand as identified by the agency through public consultation; or (H) is used to satisfy any legislative or other reporting requirements.

**Open data:** Any data that (A) is freely available in convenient and modifiable format and can be retrieved, downloaded, indexed and searched; (B) is formatted in a manner that allows for automated machine processing; (C) does not have restrictions governing use; (D) is published with the finest possible level of detail that is practicable and permitted by law; and (E) is described in enough detail so users of the data have sufficient information to understand (i) the strengths, weaknesses, analytical limitations and security requirements of the data, and (ii) how to process such data.

**Public data:** Any data collected by an executive branch agency that is permitted to be made available to the public, consistent with any and all applicable laws, rules, regulations, ordinances, resolutions, policies

or other restrictions, requirements or rights associated with the data, including, but not limited to, contractual or other legal restrictions, orders or requirements.

**Protected data**: Any data the public disclosure of which would (A) violate federal or state laws or regulations; (B) endanger the public health, safety or welfare; (C) hinder the operation of the federal, state or municipal government, including criminal and civil investigations; or (D) impose an undue financial, operational or administrative burden on the executive branch agency. "Protected data" includes any records not required to be disclosed pursuant to subsection (b) of section 1-210 of the general statutes.

**Personal data**: Any Protected data that contains personally identifiable information or protected health information.

**Private data**: Any Protected data that is subject to federal or state laws related to individual privacy or confidentiality.