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

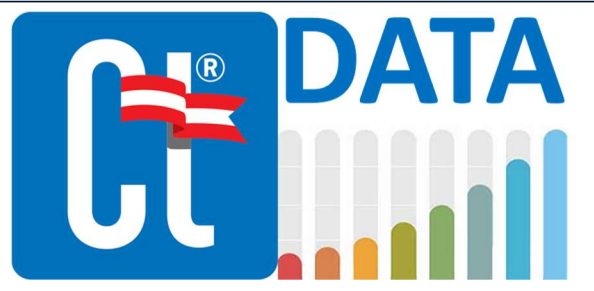Office of Policy and Management
November, 2 2020

Table of Contents

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 does not contemplate requiring agencies to dedicate additional resources to its implementation, nor does it generally require agencies to begin collecting or creating new data. Rather, it seeks to support

agency efforts and to bring those efforts into alignment to provide increased efficiency in the use of limited resources for data stewardship, use, and access.

There are many nuances that exist amongst the data collection activities of state agencies. While many agencies collect personal data on individuals that must respect the rights and confidentiality of those individuals, agencies also collect data that is inherently public and not personal in nature. This plan may not be able to specifically address every nuance that exists, and therefore is intended to be broad enough to address the many underlying challenges associated with data collection and use, while still providing 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:

<u>Preliminary draft</u> – 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.

<u>Final Plan</u> – 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
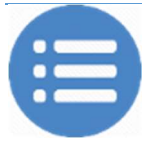
The conclusion of the first State Data Plan presents an opportunity to reflect on lessons learned. A supplemental document will report examples of progress related to the Principles and Focal Points from the inaugural State Data Plan, and related planning efforts from the past two years.

Principles

The twelve Principles below represent an ongoing framework for agencies to operate in order 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 worked to develop, automate and refine a process for daily public updates on a wide range of indicators and metrics. While it is uncertain how long the COVID-19 pandemic will impact daily life, the changes in processes, systems and management will have an impact that lasts far longer.

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
- Develop better tools to access and manage 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. The data lifecycle – from planning, through collection, management, analysis and dissemination – can promote equity to improve the ability to use data to achieve equitable outcomes. All aspects of the data lifecycle merit attention, across agency efforts to use data to inform decision-making, as disparities in housing, education, criminal justice remain.

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
- Encourage consideration of racial equity and disparate impact in the process used to access, request and publish data
- 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
- 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

Using data to inform decision-making

Whether it is called being data-informed, data-driven, evidence-based, 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:
- *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."[1]

- *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 a statewide governance structure leads to fragmented approaches to sharing data on high-priority issues which reduce the ability of the state to mobilize a response. A proliferation of data sharing agreements makes oversight difficult and reduces the ability to protect clients' data and manage risk. Flexible, durable data sharing agreements would protect clients' information and reduce the effort needed to share data.

- *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,[2] OPM identified "two approaches to coordinating existing resources – identifying established programs and identifying and monitoring proofs of concept."[3] At the same time, a 'whole-person' data view cannot be developed without attention to the ethical and equity issues inherent in developing a person-centered view without consent.

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*
  - o Continue work to implement the uniform interagency data sharing protocol, developed as part of state 2Gen efforts
  - o Continue work to expand the P20 WIN data system to cover new agencies and datasets, including social services, homelessness, financial aid and child welfare
  - o Increase transparency of agency-level efforts at performance management and work towards common standards for implementation
  - o 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*
  - o Identify opportunities for skill-building and career pathways for existing analytical staff
  - o 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*

---

[1] Legal Issues in Interagency Data Sharing, Report for Public Act 19-153, Office of Policy and Management, January 2020
[2] 2 Generational Initiative Interagency Plan, Office of Policy and Management, December 2019
[3] 2 Generational Initiative Interagency Plan, Office of Policy and Management, December 2019

- o 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
- o Align state efforts to improve delivery of services, optimize information technology 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 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:
- 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:
- Tracking agency participation in training or capacity-building efforts
- 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 metadata 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.