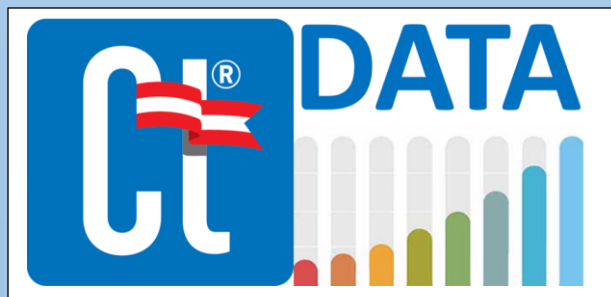


Connecting the people and processes involved with data to promote communication between, and integration of, formerly siloed data, teams, and systems within executive branch agencies.

State Data Plan

Final Draft submitted pursuant to Public Act 18-175

Office of Policy and Management
November 1, 2018



Connecticut State Data Plan

Phase 3 Draft – Submitted to the Connecticut Data Analysis and Technology Advisory Board

The State Data Plan *is not just an open data plan* but applicable to all data in the custody and control of executive branch agencies. The plan is required to be developed in accordance with [Public Act 18-175](#) . Specifically the Act requires that the state data plan shall:

- establish management and data analysis standards across executive branch agencies,
- include specific, achievable goals within the two years following 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 a 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.

An initial draft of the plan is due on or before November 1, 2018 with a final plan due December 31, 2018. The plan is required to be updated every two years thereafter.

Once final, ***information technology actions and initiatives of executive branch agencies are required to be consistent with the plan.*** Further, the Chief Data Officer is required to establish procedures for each agency to report on the agency’s progress toward achieving the goals articulated in the plan.

Table of Contents

About the Process.....	3
Introduction	4
Roles and Responsibilities:.....	5
Principles:	6
Focal Points	13
Goals	17
Glossary of Terms	19
Feedback	20

About the Process

The plan will be developed iteratively, in phases, each with an opportunity for both public and agency input.

Phase 1 – *COMPLETE*: Release draft and solicit feedback on a broad set of Principles, Focal Points, and Goals to guide the development of the plan and establish a vision for improving the management, use, and sharing of data for state agencies. Collect all feedback and use to develop Phase 2 draft.

Phase 2 - *COMPLETE*: Finalize state data plan Principles and add more specific actions under each principle that will assist in guiding agency actions. Finalize the overarching Goals and supplement with more specific measurable objectives that will advance each goal. Finalize Focal Points and supplement with additional steps to help advance implementation. Collect feedback from both the public and state agencies.

Phase 3 – *CURRENT PHASE*: Present a draft of the state data plan to the Data Analysis Technology Advisory (DATA) board established pursuant to Public Act 18-175. The DATA board shall hold a public hearing on the plan and submit any recommendations within 30 days of receiving the plan.

Final Plan: December 31, 2018

Introduction:

As we move into a knowledge-based economy, data becomes a valuable asset that the State must manage in the public trust on behalf of its residents. The effective use of data within and across state agencies is critical to enhance the ability of the agencies to formulate policies and deliver the types of services the public expects.

Agencies depend on data that may be in the custody of multiple agencies, to deliver a variety of public services. Agencies collect, use and store a wide range of personal information, such as dates of birth, social security numbers, demographic and contact details, in order to carry out their work. Examples include data that is used to verify identity, confirm eligibility and to detect and prevent fraud. While we must be vigilant in our efforts to secure and protect this data, we must also ensure that agencies proactively share data with one another and with the public.

Therefore, it is important that data are treated as State assets and managed consistently and appropriately across state government. Managing data as an asset will increase operational efficiencies, reduce costs, improve service delivery and promote better governance. While agencies must improve internal data management and use, so too must they improve the availability and accessibility of open data. Embedded within Public Act 18-175 is the presumption that data are a public asset. Proactively making public data open and accessible will enhance access to data by both sister agencies as well as the public and will enhance the public's understanding of state government, and promote public participation and civic engagement.

The purpose of this plan is to connect the people and processes involved with data to promote communication between, and integration of, formerly siloed data, teams, and systems. It seeks to promote process change, organizational realignment, and the appropriate use of technology to facilitate relationships between everyone who handles and uses data. Agencies must begin to closely connect the people who collect and prepare the data, those who analyze the data, and those who put the findings from those analyses to good use.

The plan is 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.

Roles and Responsibilities:

In addition to establishing the State Data Plan, Public Act 18-175 creates 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.”

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;
- 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;
- 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.

Principles:



The Principles represent a framework under which agencies should operate in order to mature and enhance their management, use, sharing, and analysis of data. Each Principle includes a series of practices that agencies should begin to adopt and implement during the two year period that this plan covers.

1. Appreciate that the data we collect require a considerable investment of resources and have value beyond the purpose for which they are collected.

Agency Practices

- a. Manage and treat data as a strategic asset by defining strategic priorities for the management and delivery of data with an emphasis on accuracy, completeness, and timeliness.
- b. Maintain an inventory of high value data in accordance with Public Act 18-175.
- c. Establish strategies to ensure the value of the data collected is maximized through the analysis and sharing of data, and by identifying new opportunities to leverage data.
- d. Ensure that investments in new or upgraded data collection systems provide bulk access to raw data, access through Application Programming Interfaces (API's), and that the data are owned by the state.

Support from the Chief Data Officer:

- Provide data inventory template and guidance.

2. Perform continuous data quality and analytics improvement to ensure the value of data is protected and maximized.

Agency Practices

- a. Establish feedback loops amongst various components of the data pipeline to allow for improvement of data at the source.
- b. Implement Lean and Agile practices in consultation with the Chief Data Officer and Process Improvement units of OPM.
- c. Measure the results of data analysis and analytics to ensure such activities are driving actionable insights and creating value.

Support from the Chief Data Officer:

- Development of a Lean/Agile framework and guidance.

- Coaching and consultation in collaboration with the state's Process Improvement Program Director.

3. Ensure standardized data governance in order to protect data and improve its quality and utility.

Agency Practices

- Establish governance practices that rely on best practices that mitigate risks to data and ensure data are made available for efficient use.
- Ensure the ability to exercise authority, control, and shared decision making (planning, monitoring, and enforcing) over the management of data assets.
- Define roles and responsibilities related to data quality, access, and decision making that are assigned and documented.
- Outline a policy to structure appropriate data controls, including access, availability, and methods to ensure quality.

Support from the Chief Data Officer:

- Consultation and guidance to agencies related to data governance best practices.
- Templates to aid in formulating data governance charters, structures, etc.
- Resource library of documents and material related to best practices.
- Maturity scale for data governance that agencies can use to plan for enhancing their data governance efforts.

4. Protect individual privacy and maintain promised confidentiality using effective data stewardship and governance, and by maintaining modern data security practices and technology

Agency Practices

- Establish a suppression policy for use in aggregating private data. This will help agencies establish a standard to follow, and allow for flexibility to modify if a particular type of data require different privacy rules.
- Ensure awareness of the scope of data which may be at risk and what data the agencies collect.
- Complete a data classification assessment for all data identified in the agency high value data inventory
- Ensure that policies and procedures with respect to the processing of personal data (that which includes personally identifiable information or protected health information) cover:

- i. Personal data held by the agency or which is made available to other agencies are necessary for the purposes of fulfilling a legal or regulatory requirement, or for delivering a public service;
 - ii. Procedures for an individual to obtain more information on or gain access to the personal data which the individual has previously provided to the agency, including the use and disclosure of such data by the agency; and
 - iii. Contact details of the individual to whom enquiries or feedback on personal data can be forwarded.
- e. Ensure that data is collected by lawful and fair means, and is limited to that which is necessary to fulfil its statutory or business requirements.

Support from the Chief Data Officer

- Provide data Classification Assessment template.
- Provide model policies and procedures.
- Provide model suppression policy and standards.

- 5. Employ ethical standards in the use and analysis of data to avoid intrusion into the lives of Connecticut residents.

Agency Practices

- a. Recognize that data are a public asset and should be used to promote the public good.
- b. Evaluate, recognize and prevent any possible unintended consequences to individuals, their privacy, their safety, and their opportunities.
- c. Protect data and data use from institutional and political bias.

Support from the Chief Data Officer

- Provide an ethics assessment and risk management tool including guidance.

- 6. Promote a culture of continuous and collaborative learning with data and about data.

Agency Practices

- a. Increase the level of data literacy within agencies through training and education.
- b. Perform data audits to understand what data is collected, how it's collected, the technical system of collection, who uses the data and for what purpose.
- c. Analyze and understand the extent to which bias exists in the data, models, methods, and analysis and the impact it has on equity, stability, and opportunity.
- d. Establish mechanisms and procedures to share code, software, services, and tools.

- e. Establish feedback loops and opportunities for public participation related to the value, accuracy, and utility of data.
- f. Identify and reduce barriers related to the ability to interpret and analyze data.

Support from the Chief Data Officer:

- Establish a Data and Analytics Center of Engagement.
 - The Center shall advise and consult agencies in developing efficient data and analysis processes, skill development, and technology solutions.
 - Includes OPM IT Policy Bureau, state Process Improvement Program, and the Department of Administrative Service Bureau of Enterprise Systems and Technology;
- Establish a catalog of training resources
- Identify and share best practices from other States.

- 7. Create, acquire, use, and disseminate data deliberately and thoughtfully; considering legal and regulatory constraints, quality, consistency, privacy, value, reuse, and interoperability from the start.

Agency Practices

- a. Establish plans for the collection of data to ensure (1) the data are maintained throughout their useful lifecycle, (2) adhere to existing standards, and (3) are well documented including metadata and data dictionaries.
- b. Standardize common data elements, when investing in new or upgraded systems, to ensure data are easily shared, integrated, and analyzed.
- c. Ensure that definitions and formats of data that are collected in existing systems are known and how they correlate to existing or adopted standards is documented.

Support from the Chief Data Officer:

- Establish recommended standards for common elements.
- Establish a metadata and data dictionary repository.

- 8. Coordinate and prioritize data needs and uses, utilize data from multiple sources, and acquire new data only when necessary.

Agency Practices

- a. Consult with agencies and the Chief Data Officer prior to initiating data collection efforts to avoid redundant or duplicative efforts.
- b. Identify gaps in data, or data that the State does not collect, and prioritize acquisitions that can aid in enhancing the effectiveness and efficiency of state policies and programs.

- c. Leverage multiple sources of data for analysis in order to ensure a comprehensive view of an issue, policy, or program outcome.

Support from the Chief Data Officer:

- Facilitate interagency coordination and collaboration through round table discussions and other communication methods.

9. Manage a data asset one time and use it for multiple purposes, to the extent possible within legal and regulatory constraints.

Agency Practices

- a. Establish access to data sources via Application Programming Interfaces (APIs) when investing in new or upgraded data systems.
- b. Share the data collected or generated by an agency with other agencies, within the bounds permissible by applicable laws and the principles of data privacy , to allow the state to achieve the objectives of:
 - i. delivering customer-centric services;
 - ii. developing appropriate and meaningful policy;
 - iii. achieving quality policy formulation;
 - iv. facilitating useful analysis and research.

Support from the Chief Data Officer:

- Provide advice and consultation.
- In consultation with DAS/BEST evaluate the feasibility of establishing an Enterprise Service Bus or comparable means to enable API accessibility.
- Access to the enterprise data integration hub which currently is available through DAS/BEST.

10. Improve data sharing and access with ongoing input from users and other stakeholders.

Agency Practices

- a. Engage partner agencies and other data users to streamline data sharing processes and improve the quality and utility of agency data.
- b. Identify and prioritize high-impact, high value data assets to prioritize in the formulation of agency Open Data Access Plans.
 - i. Engage external partners including academia, non-profits, and other stakeholders when releasing Open Data.
- c. Measure and track usage of open data by the public, including raw downloads and by examples of valuable usage

Support from the Chief Data Officer:

- Facilitate interagency coordination through round table discussions and other communication methods
- Provide tools and guidance to support assessing impact and the value of open datasets.
- Provide data and tools to assist agencies in accurately measuring open data use.

11. Create clear and predictable pathways for data sharing which are necessary for effective data use and sharing.

Agency Practices

- a. Develop standard protocols and procedures for data sharing.
- b. Establish documented processes, which outline both the technical and legal aspects necessary for data sharing.
- c. Provide examples or use cases that involve legal and permissible uses of data.
- d. Provide templates or “boilerplate” language for data sharing Memoranda of Agreement and/or Data Use Agreements.

Support from the Chief Data Officer:

- Catalog existing data sharing agreements
- Catalog best practices and use cases from other states.

12. Embrace openness, transparency, and accountability.

Agency Practices

- a. Proactively release public data as open data such that the data is robust and well-documented in order to fulfill the greatest number of uses
 - i. Provide open data at the finest level of geographic and demographic granularity as possible
- b. Aggregate private data in consistent and meaningful ways, to enable policy makers to make better decisions.
- c. Leverage open source tools and technology, share agency code and tools under open licenses in order to capitalize on shared knowledge and support others.
- d. Ensure adherence to the state’s freedom of information and open records laws, and presume data are public unless otherwise protected by state or federal law, in accordance with PA 18-175.

Support from the Chief Data Officer:

- Provide an online repository for the publication of open data (Open Data Portal: <https://data.ct.gov>)
- Develop dataset templates and guidance for aggregating data

- In consultation with OPM IT Policy Bureau and the Department of Enterprise Service and Technology, establish an open source policy related to data and data analytics software and tools.

Focal Points



The Focal Points represent areas where state agencies should emphasize the sharing, integration, and availability of data. The Focal Points are not necessarily intended to represent the most pressing policy areas but may represent important issues 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.

Education & Workforce Development:

- Build upon and scale the Preschool through 20 and Workforce Information Network (P20 WIN) longitudinal data system to:
 - Include occupational and professional licensing data.
 - Include early childhood data.
 - Establish more sustainable and efficient mechanisms to integrate and match data across systems,
 - Advance the use of more robust and timely performance data throughout the P-20 education and workforce development systems to support improvement efforts benefitting children, families and Connecticut's economic competitiveness
 - Proactively identify and produce an annual de-identified integrated dataset for research and evaluation purposes, with elements identified through stakeholder engagement, for secure sharing with authorized researchers in accordance with P20 WIN's existing data request processes.
 - Identify through stakeholder engagement, and publish as open data, aggregate datasets produced by P20 WIN.
- Through stakeholder engagement, identify and prioritize high-impact, high-value public datasets for release as open data.
- Identify and pursue mechanisms to integrate data related to employment and educational outcomes including public assistance recipients, income status, and other barriers to employment, education, and training.

Healthcare and Service Delivery:

- Support efforts to expand the secure master person index among various health and human services agencies necessary for data integration at both an operational or case management level as well as to support anonymized integrated datasets for research and evaluation purposes.
 - Focus efforts on agencies that can leverage federal funds for this purpose
- Through stakeholder engagement, identify and prioritize high-impact, high-value public datasets for release as open data

- Ensure such data are disaggregated, to the extent permissible by law, by town and by standard demographic characteristics including race, ethnic, gender, etc.
- Prioritize data that can aid in identifying gaps in services or disparities in outcomes.
- Integrate data that can drive healthcare decision making, policy, and service delivery by establishing solutions to allow data from the various state agencies to be accessed, analyzed, and visualized.
- Identify opportunities to acquire, integrate, or use new data sources not already collected by state agencies

Fraud, Waste, and Abuse:

- Scale the analytical capabilities currently in place to onboard additional agencies and data sources to advance the state’s efforts to detect fraud, waste, and abuse associated with state supported assistance programs.
- Identify additional programs and data assets necessary to support research and data analysis.
 - Establish a strategy to address any legal or financial barriers to participation
- Identify and consider the use of data assets not currently held by the state.

Substance Abuse and Addiction:

- Enhance both the coordination of data analytics around substance abuse, addiction and overdoses and facilitate timelier sharing of data among local, state, and federal government agencies.
- Align and coordinate data related work groups of various task forces and individual agency efforts that currently overlap.
- Identify and catalog key agency data assets necessary to support research and data analysis.
 - Identify legal and other barriers to data sharing that may exist.
- Identify agency staff with data analysis skills, especially geospatial analysis, which can support data integration and analysis efforts.
- Identify and catalog key federal or local data assets not currently collected that can support research and data analysis efforts.
- Through stakeholder engagement, identify and prioritize high-impact, high-value public datasets for release as open data.

Resilience:

- Leverage data to support strengthening the state’s resiliency to extreme weather events including hurricanes, flooding, extreme heat, and slow onset events such as sea-level rise.
- Identify and catalog key agency data assets necessary to support the state’s ability to plan for, prepare for, respond to, and recover from extreme weather related events.
- Through stakeholder engagement, identify and prioritize high-impact, high-value public datasets for release as open data and give priority in Open Data Access Plans.

- Coordinate activities with the State Agencies Fostering Resilience (SAFR) and the Division of Emergency Management and Homeland Security’s Emergency Management Support Function 5 (ESF 5) data Integration working group.
- Identify and catalog key federal, local data, or other assets not currently collected that can support research and data analysis efforts.
- Establish a platform to allow data from the various entities to be accessed, analyzed, and visualized, with an emphasis on geospatial capabilities.

Traffic Safety:

- Support existing efforts based on the National Governors Association learning lab focused on state strategies to improve statewide data use and sharing to reduce traffic fatalities and injuries.
- Continue efforts to maintain and update a comprehensive traffic records data inventory and expand it to include data dictionaries and useful metadata.
- Identify legal and other barriers to data sharing that may exist, and establish data sharing agreements as necessary.
- Through stakeholder engagement, identify and prioritize high-impact, high-value public datasets for release as open data.
- Establish a mechanism to “link” data across systems and agencies.

Economic Development:

- Undertake efforts to support economic development in the state by making additional data to support the state’s business development efforts available to key stakeholders.
- Integrate data from various state agencies to support streamlined and efficient business processes.
- Identify and catalog key agency data assets necessary to support the state’s ability to enhance economic development efforts.
- Through stakeholder engagement, identify and prioritize high-impact, high-value public datasets for release as open data to support Connecticut business and economic development communities.
- Identify and catalog key agency data assets which can be integrated in order to streamline agency business processes such as permitting, licensing, registration, and reporting.

Land Use:

- Support the efforts of Connecticut’s planning community to plan for and adapt to the impacts of change not only to related to Connecticut’s physical landscape but also economic development, transportation, social equity and many other disciplines.
- Through stakeholder engagement, identify and prioritize high-impact, high-value public datasets for release as open data.
- Establish a common geospatial data framework which could be used by state and local governments.

- Establish strategies to develop, update, and maintain a statewide property (parcel) dataset.
 - Update the statewide parcel (cadastral) standard in consultation with the planning and geospatial communities
 - Establish strategies to develop, update, and maintain a statewide address point dataset (anonymous physical location of addresses with geographic coordinates)
 - Establish strategies to develop, formalize, and maintain neighborhood shapefiles for each municipality.
-
- Establish a platform to allow data from the various entities to be accessed, analyzed, and visualized, with an emphasis on geospatial capabilities.
 - Establish connections between the planning community and Connecticut's civic technology community to enhance the availability and accessibility of data and tools.
 - Establish strategies to ensure the continued acquisition of high-resolution aerial imagery on a regular basis, and ensure its availability.

Goals



The overarching purpose of the State Data Plan is to facilitate the sharing and use of data amongst executive branch agencies in order to enhance the efficiency and effectiveness of state policies and programs. The following goals, when pursued collectively amongst agencies, will advance this objective.

- 1.) Reduce the effort necessary to acquire and prepare data allowing for more resources to analyze data and to improve outcomes for Connecticut residents
 - Identify, and where possible, remove data sharing barriers between state agencies.
 - Improve the quality of agency data such that it is rapidly suitable for analysis
 - Enhance self-service access to data and analytical tools.

- 2.) Create more impactful and predictable open data that's useful to both the public and state agencies in order to enhance data sharing, drive efficiencies, increase civic engagement, and create new insight.
 - Increase the percent of open data that meets documentation and metadata standards
 - Establish, maintain, and adhere to Open Data Access Plans
 - Measure and track open data use including downloads, views, and use cases.

- 3.) Improve knowledge and understanding of data that is held by state agencies
 - Complete, maintain, and continuously improve upon agency high value data inventories.
 - Examine the reliability and statistical validity of agency data.
 - Increase availability of and access to agency metadata and data dictionaries.

- 4.) Increase standardization and integration of data and systems
 - Identify and establish standards for common data elements across agencies with priority given to:
 - Standards for locational data such as addresses, towns, and zip codes.
 - Standards for demographic data such as race, ethnicity and language.
 - Improve the documentation of existing data and systems such that the relationship to existing or adopted standards is known.
 - Increase the number of systems and data sources accessible via Application Programming Interfaces (APIs).
 - Expand use and adoption of the State's enterprise data integration hub.

- 5.) Improve the knowledge and skills of state employees related to data stewardship and analysis
- Increase the number of state employees with basic data literacy training.
 - Identify and participate in more advanced data and analytics trainings.
 - Identify skills necessary for data and analytics in the state workforce, identify any current job classifications that may align with those skills.
 - Establish a strategy to fill the skills gap by determining the feasibility of new job classifications, or identifying partnerships to supplement existing capabilities.

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.

Executive branch agency: Any agency listed in section 4-38c of the general statutes, except the Board of Regents for Higher Education.

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.

Feedback

The deadline to submit public comments or feedback is November 15, 2018

- A public hearing will be held by the Connecticut Data Analysis Technology Advisory Board on November 15 at the Legislative Office Building in Hartford.
- Those unable to appear in person at the public hearing may submit your comments [via email](#).
- Additionally, the Draft is available as a [Google Doc](#) with the option to comment directly.