

DATA Board August 2022 Meeting

Connecticut Data Analysis and Technology Advisory Board
August 4, 2022 Meeting



Agenda

- Welcome and Introductions
- Updates
- Open Data Portal
- Interagency Data Sharing
- Geospatial Information Systems
- Planning for 2023 – 2024 State Data Plan
- Next steps



Welcome and Introductions

DATA Board role, powers and duties

CGS Sec. 2-79e:

The board shall have the following powers and duties:

- (1) To advise the executive, legislative and judicial branches of government and municipalities concerning data policy, including, but not limited to, best practices in the public, private and academic sectors for data analysis, management, storage, security, privacy and visualization and the use of data to grow the economy;
- (2) to advise the Office of Policy and Management regarding the online repository established under section 4-67p;
- (3) to issue reports and recommendations in accordance with section 11-4a;
- (4) upon the request of at least two members of the board, to request any agency data officer or agency head to appear before the board to answer questions;
- (5) to request from any executive department, board, commission or other agency of the state such assistance and data as necessary and available to carry out the purposes of this section;
- (6) to make recommendations to the legislative leaders and the directors of the Offices of Fiscal Analysis and Legislative Research regarding data analysis skills and related expertise that the leaders and said offices may seek to cultivate among their staff through training or as a consideration when hiring staff; and
- (7) to establish bylaws to govern its procedures.

Chief Data Officer and State Data Plan

CGS Sec. 4-67p: “The Chief Data Officer shall be responsible for:

- (1) directing executive branch agencies on the use and management of data to enhance the efficiency and effectiveness of state programs and policies,
- (2) facilitating the sharing and use of executive branch agency data (A) between executive branch agencies, and (B) with the public,
- (3) coordinating data analytics and transparency master planning for executive branch agencies, and
- (4) creating the state data plan in accordance with subsection (c) of this section.”

Data and Policy Analytics at OPM

The [Data and Policy Analytics Unit \(DAPA\)](#) was established to serve as a resource for data analysis and data-related projects at the Office of Policy and Management.

The DAPA unit directs and manages activities related to the collection, analysis, sharing, coordination and dissemination of data and is responsible for the development and implementation of the [State Data Plan](#), the [Open Data Portal](#), and [P20 WIN](#), the State's integrated data system, and development of the State's GIS Office for geospatial data.

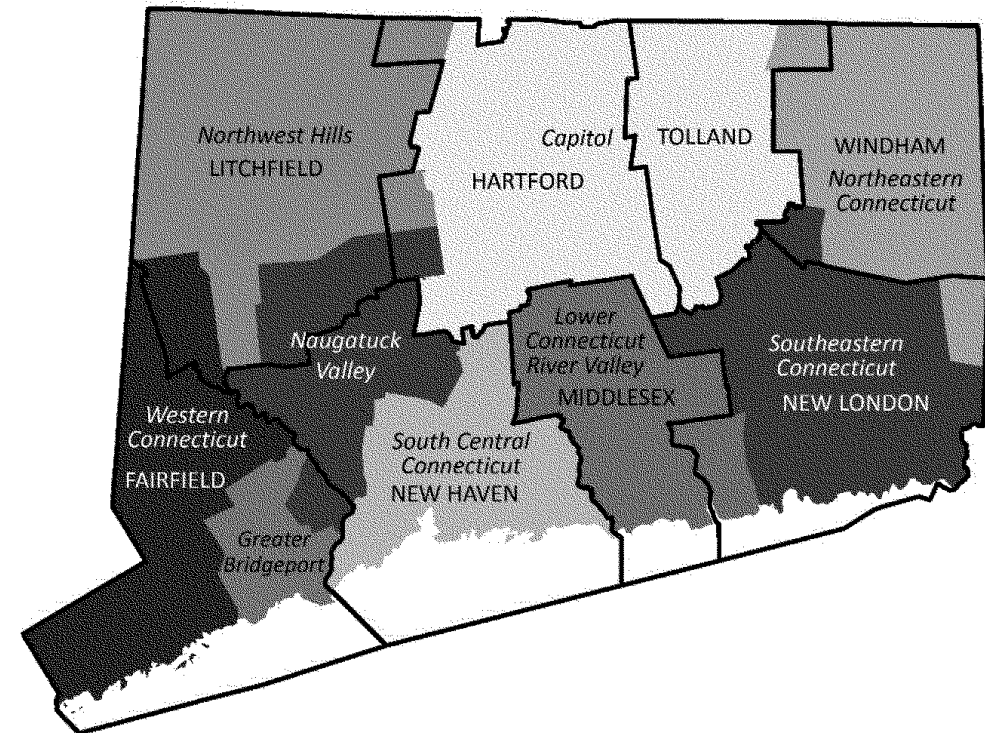
The use of data across agencies facilitates a system-wide perspective and is a critical component of Governor Lamont's statewide performance and results agenda to implement a data-driven and informed policy agenda in state government that is user-friendly, cost-effective, data-informed, and results-driven.

Updates

Census Updates

Census issued final Federal Register Notice for transition to county-equivalents on June 6, 2022

Figure 1. Relationship between Counties and Planning Regions. County boundaries are shown as thick lines; names are labeled in standard font in all caps. The planning regions are shaded; names are in italics.



Legislative changes

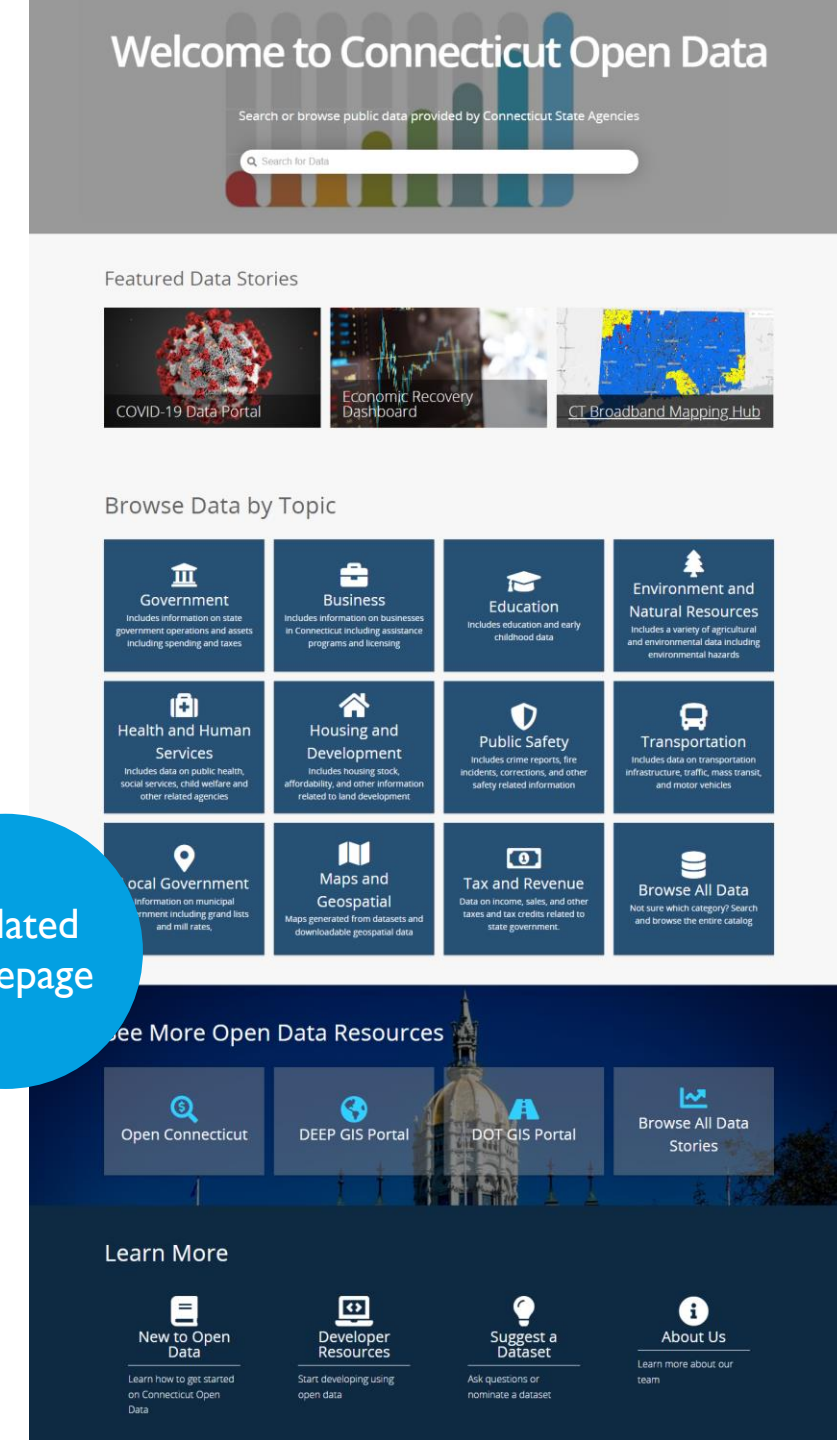
- PA 21-2, Sections 284, 298, 303, 305, effective July 1, 2021:
 - Amends 10a-57g to expand P20 WIN's purpose to also include informing policy and practice for **education, workforce, and supportive services**
 - Reconstitutes the P20 WIN executive board: includes leader of each participating agency, OVVS, and OPM, or his or her designee
 - New data collections on credentials, private occupational schools, enhanced wage records
- PA 21-2, Sections 78 - 79:
 - Establishes GIS Office, GIO, GIS Advisory Council (more later)
- PA 21-35, Section 11:
 - Requires standard categories for race, ethnicity and language for healthcare providers and health and human services agencies, administered by Office of Health Strategy



Open Data Portal

Open Data Portal Background

- Launched in 2014, the Connecticut Open Data Portal is the state's repository for open data.
- Managed by the Data and Policy Analytics team at the Office of Policy and Management (OPM).
- Our team works with staff at other state agencies to publish their own data on the portal.
- Visit the portal at data.ct.gov.



Making state data available as **open data**

[Section 4-67o of the Connecticut General Statutes](#) defines open data as any data that:

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

Open data snapshot

What's on the Portal?

592
open datasets

58
maps

144
charts

37
data stories

To what extent are users accessing the data?

11.4 million
asset visits since 2014

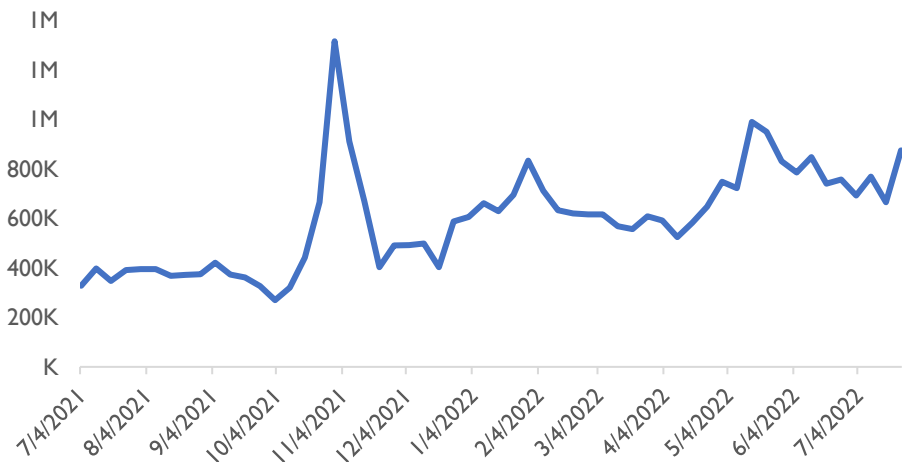
4.3 million
downloads since 2014

How many users are active?

102
data editors/publishers active in 2021-2022

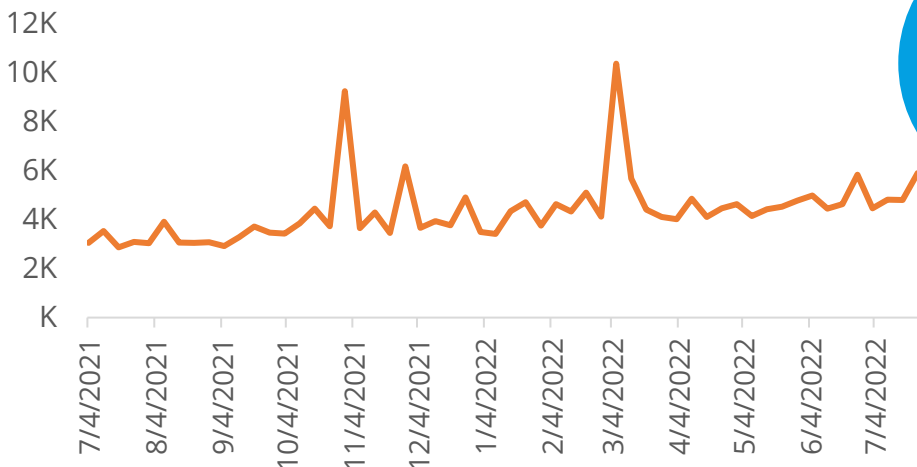
Open data trends, July 2021 - July 2022

Weekly API reads



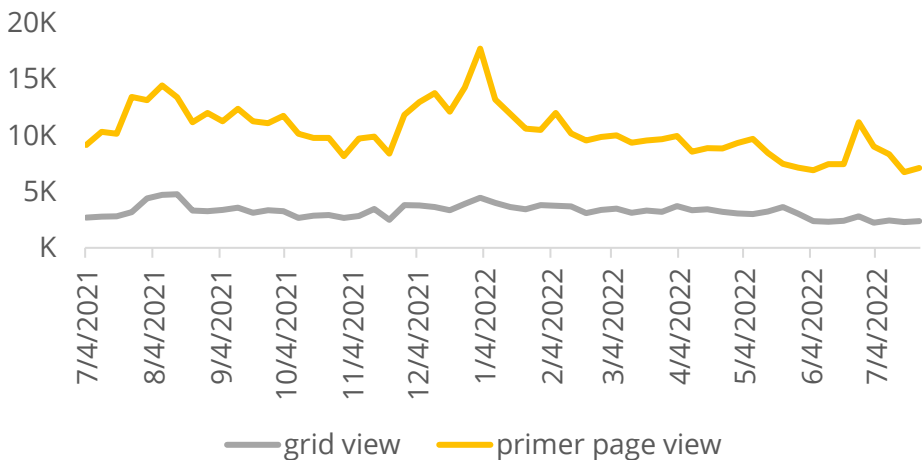
596,442
average
weekly API
calls

Weekly downloads



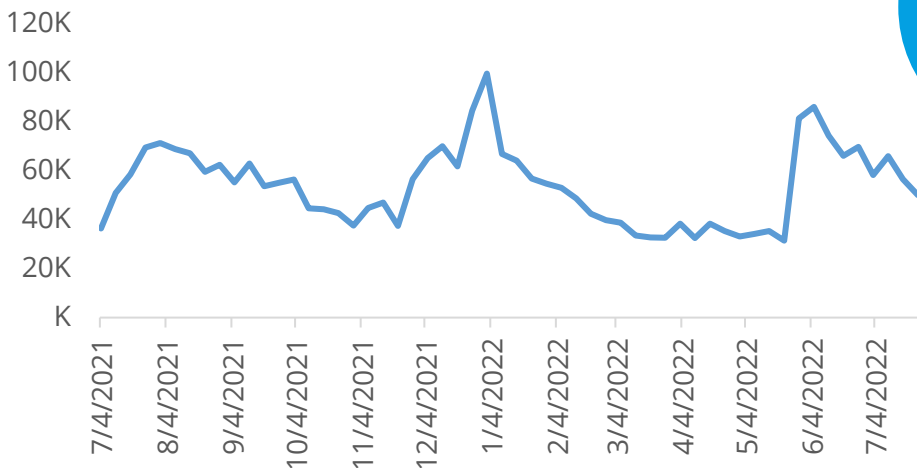
4,354
average
weekly
downloads

Weekly dataset views



13,700
average
weekly views

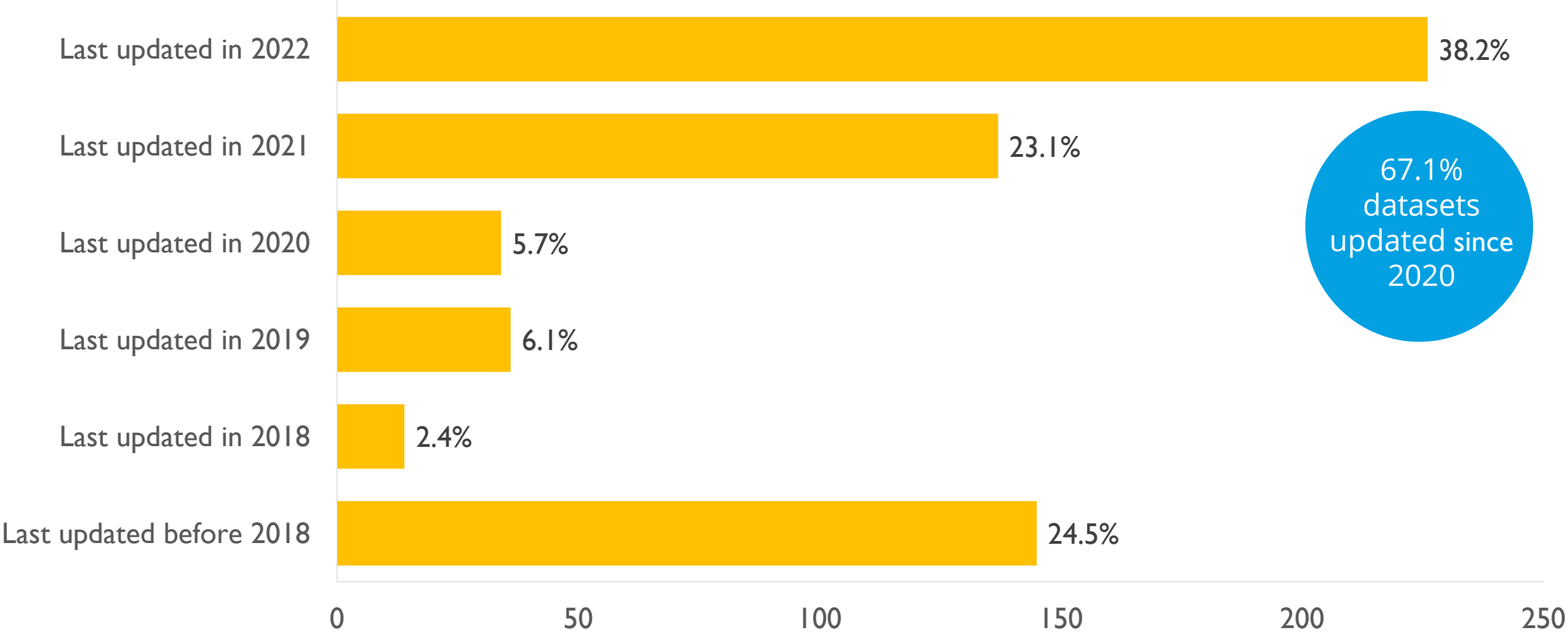
Weekly data story views



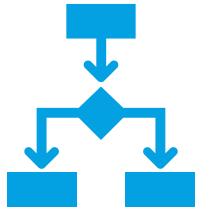
53,827
average
weekly views

Data freshness

Datasets by last date updated



Current initiatives



1. Developing **processes and standards** to improve data governance



2. **Cleaning up assets** that are no longer being maintained



3. Highlighting data through **data stories**

New processes and standards

The DAPA team is working to develop new guidance for open data publishers to improve the quality of the data on the Portal. New guidelines include:

- [Data retirement process](#)
- [Data publication guidelines](#)
- [Automation guidelines](#)
- Data visualization guidelines (in progress)
- Data story style guide (in progress)

Open data cleanup

Over the last year, the DAPA team has been working with agencies to clean up assets on the Portal that are no longer being maintained by:

- **Reviewing** the assets on the portal attributed to each agency,
- Working with agencies to **update** old assets,
- **Consolidating** assets that are published as new assets every year into a single multi-year dataset
- **Retiring** old assets that are no longer being updated
- **Automating** updates to the portal where appropriate

Open data cleanup **progress**

- 254 assets retired since June 2021
 - 65% of retired assets had not been updated since before 2018
 - Following the process laid out in the [Data Retirement Guidelines](#) document
 - Retired assets are logged in this dataset on the Portal: [Assets retired from data.ct.gov](#)
- Process has contributed to improved data quality, ensuring that more data is up-to-date
 - 67% datasets have now been updated since 2020
- Ongoing process to retire, consolidate, and update data

Data stories

- Data stories are pages on the Open Data Portal that can be used to present data visualizations alongside explanatory text
- Useful for providing context and increasing access to data published on the Portal
- Our first data stories were published in 2020 to provide greater access to data on COVID-19

[Developers](#)[Suggest a Dataset](#)[Help](#)[About](#)[Sign In](#)

Connecticut Economic Recovery Dashboard

[Home](#)[Small Business](#)[Social Services](#)[Housing](#)[Employment](#)

Access to the
underlying
data

This dashboard displays various indicators on the economic recovery in Connecticut since the start of the COVID-19 pandemic. To see the source data for each chart, click the link in the citation under the chart, or browse all the data used in this dashboard [here](#).

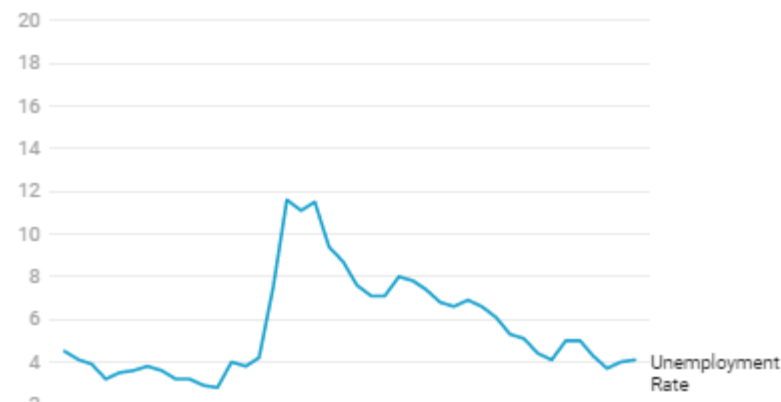
Cross-State Unemployment Rate

Click on the buttons below to see data from different time periods:

[Average 2009](#)[Average 2019](#)[Most Recent Month Available](#)

State name	Period name	Unemployment Rate
Connecticut	May 2022	4.0
Massachusetts	May 2022	3.4
New Jersey	May 2022	3.4
New York	May 2022	4.1
Rhode Island	May 2022	2.7

Unemployment Rate: Connecticut

[Connecticut](#)[New Jersey](#)[New York](#)[Massachusetts](#)[Rhode Island](#)

Text that
provides
context

Tables, charts,
other
visualizations

Recent data stories

Data stories published in the last year have covered various topics including:

- Public health
 - [School Immunization Survey Data](#)
 - [Childhood Lead Poisoning Surveillance Report](#)
 - [Childcare COVID-19 Surveillance](#)
- Education
 - [Roberta Willis Scholarship Program Analysis](#)
- Government and sustainability
 - [GreenerGov CT Dashboard](#)

Next steps for the Open Data Portal

1

Continue cleanup efforts

Work with agencies to retire, consolidate, and update datasets to improve usability of the Portal

Implement system to track data freshness and identify datasets in need of updating

2

Processes and standards

Finalize guidance documents in progress on data visualization and data stories

Improve presentation of existing guidance documents for greater accessibility by users

3

Metadata improvements

Review metadata for completeness

Update metadata standard

Implement tagging strategy to improve search functionality

4

Collaboration with CT Data Collaborative

Publish data from U.S. Census Bureau and from state agencies that the CT Data Collaborative has published on CKAN website previously

Interagency Data Sharing

OVERVIEW

- [P20 WIN](#) is Connecticut's state longitudinal data system and is the mechanism by which data from multiple agencies are matched to address critical policy questions. P20 WIN informs sound policies and practice through secure sharing of longitudinal data across participating agencies to ensure that individuals successfully navigate supportive services and educational pathways into the workforce.
- P20 WIN is a federated data system that has been operational since 2014. P20 WIN is used to answer policy questions, fulfill federal and state reporting requirements; support program review; inform school districts of postsecondary outcomes; provide employment and wage outcome data; and support research and analysis on a variety of topics.

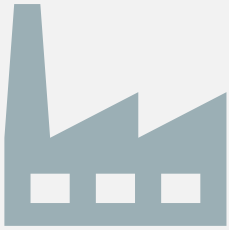




MEMBERSHIP

- **Participating Agencies**
 - Department of Labor (DOL) – Also the Data Integration Hub
 - State Department of Education (SDE)
 - Office of Early Childhood (OEC)
 - CT State Colleges and Universities (CSCU)
 - University of Connecticut (Uconn)
 - Department of Social Services (DSS)
 - Department of Children and Families (DCF)
 - Office of Higher Education (OHE)
 - CT Conference of Independent Colleges (CCIC)
 - CT Coalition to End Homelessness (CCEH)

ADMINISTRATION



Operating Group

P20 WIN is administered by the Office of Policy and Management (OPM) as the Operating Group, which provides program management to support the continued operation and improvement of P20 WIN as a resource for the participating agencies and the State.



Data Integration Hub

The Data Integration Hub for P20 WIN is based at the Department of Labor, where all matching for approved requests happens.

GOVERNANCE STRUCTURE

- P20 WIN has an **Executive Board**, a **Data Governing Board** and **Data Stewards** that provide the structure to create and implement the necessary policies and procedures for a multi-agency Data Sharing system to address broad policy questions and state needs.
 - Respond to data requests.
 - Determine authorized Users and access rights.
 - Develop and maintain a cross-agency Data Dictionary
 - Add, Suspend and Readmit Participating Agencies
 - Responsible for P20WIN Policies, Staffing and Financial Support
- **Data Stewards** are identified by each Participating Agency to maintain and create datasets for approved data requests.
- The **Resident Advisory Board** will provide accountability and transparency for data requests that seek to use data from marginalized communities. RAB members will have lived experience in poverty and have been or are currently users of state services.

THE LEARNING AGENDA PROCESS



LEARNING
AGENDA
DEVELOPMENT

LEARNING AGENDA

The **P20 WIN Learning Agenda** reflects the priority issues and burning questions of the State of Connecticut and P20 WIN Participating Agencies. The topics and questions in the Learning Agenda are used to prioritize and inform data requests to P20 WIN.

- **College and Career Success:** to support placement decisions at colleges and universities
- **Student Readiness:** to analyze transitions between early childhood and K-12 schools, to include social services, child welfare, housing, family life and adult education
- **Financial Aid:** the dynamics and outcomes for state financial aid grant recipients
- **Workforce Training:** measuring the net impact and return on investment for public workforce training programs
- **Overcoming Barriers to Success:** using data to help individuals who face barriers to success due to factors like homelessness or engagement with the child welfare system

LEGAL AGREEMENTS

P20 WIN uses a series of enterprise agreements to facilitate the data sharing process across Participating Agencies. The agreements include an Enterprise Memorandum of Understanding and Data Sharing Agreements.

Enterprise Memorandum of Understanding (eMOU)

- Signed by executive leadership from all Participating Agencies of P20 WIN
- Facilitates the operation of the system
- Establishes the governance structure and request management process
- Sets security and privacy standards

Data Sharing Agreements (DSA)

- A formal document signed by the Participating Agencies, Data Integration Hub and the data recipient for an approved data request.
- Clearly details what data are being shared and how the data can be used.
- Includes provisions concerning access and dissemination, including:
 - Requirements to access specific data
 - Confidentiality and disclaimers
 - Data destruction
 - Time limits

Data sharing playbook: key stages of the process



Data Requester



Key Step



Data Owner

Locate data

Identify the Data Owner

Connecticut's High Value Data Inventory

Public Data Dictionaries

Find Data Owner's data sharing process

Record data in CT's High Value Data Inventory

Create Public Data Dictionaries

Create a data sharing process

Build a data request

Make the case for data sharing

Specify the parameters of the request

Define the scope and timeframe

Making and receiving a data sharing request

Ensure key questions are answered

Safeguard your data

Identifying type of legal agreement needed

Data sharing playbook: key stages of the process



Data Requester



Key Step



Data Owner

Sign the agreement

Sign the agreement



Creating a data sharing agreement



Sign the agreement

Secure data

Establish Credibility
working with data



Establish privacy and security
processes

Understand the legal and
regulatory environment



Safeguard your data

Send and receive data

Receive data



Transferring data through
appropriate channels



Securely share data

(If needed) Deidentify or
Aggregate data



Geospatial Information Systems



GIS Office & GIO

OPM's Geographic Information Systems (GIS) Office was established in 2022 following passage of Public Act 21-2 during the 2021 June Special Session. It is directed by a Geographic Information Officer (GIO) and resides within the Data and Policy Analytics Unit of OPM.

The GIS Office is responsible for:

Coordinating the collection, compilation and dissemination of GIS data across the state, including from and to state agencies, regional councils of governments, municipalities and other constituencies;

Managing a publicly accessible geospatial data clearinghouse;

Using GIS to support economic development efforts in the state;

Provide training and outreach on the use of GIS;

Administering a statewide orthoimagery and lidar program;

Adopting geospatial data standards, guidelines, and procedures;

Performing technical data processing to aggregate and organize existing datasets and create new datasets; and

Develop broadband data and mapping in accordance with Public Act 21-159.



GIS Office Staff

Geographic Information Officer (GIO)

Started January of 2022.

Alfredo
Herrera

9+ years of experience as a GIS professional.

Most recently spent 5 years transforming the GIS program at the City of New Haven, CT.

GIS
Office
Staff

David Lukens – Broadband Mapping
Coordinator

GIS Analyst – To be Hired Soon (Interview
Process is ongoing)



Important Milestones

First two meetings of the GIS Advisory Council (meetings occur every 2 months)

3 out of 4 Advisory Council workgroups have started work.

Launched Broadband Mapping Hub Website <https://broadbandmaps.ct.gov>

Completed MOU with DOT to accelerate completion of updated administrative boundaries.

GIS Office Priority Topics





Broadband Mapping

Published broadbandmaps.ct.gov

Completed Tasks

ISP data submissions received from April to May

Initial cleaning and processing to request additional information



Broadband Mapping

To-Do

Cleaning and alignment of address-level data to existing data layers (e911, etc.)

Processing and analysis

Unserved and underserved addresses

Access and adoption variance

Affordability analysis

Third-party verification

CCM survey

Broadband hub address check



Aerial Imagery Data Acquisition

Workgroup meeting monthly.

Eric Lindquist selected as Project Lead.

RFP currently being written, slated to be released mid-September.

Two imagery and LiDAR captures 3 years apart.




Geospatial Open Data Clearinghouse

Workgroup meeting bi-weekly.

Reviewing other states' clearinghouses and synthesizing results.

Gathering GIS community needs

Working through technology procedures, changes, and challenges.



Statewide Addressing and Geocoding

Improving statewide addressing datasets to increase their utility.

Assessed Dataset Challenges.

Using utilities' service address info as basis for improvement of this dataset.

Collaborating with DESPP and DOT to ensure maintenance workflow is established.

Developing plan for sustainable validation and correction of submitted changes.



Parcel Data Collection and Standardization

Version 5 of CAMA Schema was implemented this year

Contains 136 standardized field names

Expecting approximately 90% of town assessors to utilize the standard report this year; hoping to increase that percentage every year

Will enable easier joining of CAMA data to parcel files. Ability to utilize unique identifier scheme to crosswalk assessment data to parcel files

CAMA schema will become default parcel file attribution

OPM hiring research analyst to help clean/process the data in the future

OPM collects CAMA and parcel data from the COGs in May-June, then makes available for download

Metadata will be provided (collected from the towns)

Upon using the data, report any issues or requested improvements to Eric Lindquist eric.k.lindquist@ct.gov. Eric is compiling a list of issues to resolve with the next version of the standard CAMA schema.



Parcel Drafting Standards

Creating a guidebook for GIS professionals to encourage a standardized approach to parcel boundary digitization and presentation.

Workgroup meeting bi-monthly.

Work assignments to workgroup members have been assigned and work started.

2023 – 2024 State Data Plan

What is the state data plan?

According to [section 4-67p](#) of the Connecticut General Statutes:

“The state data plan shall

- (1) **establish management and data analysis standards** across all executive branch agencies,
- (2) include specific, achievable goals within the two years following adoption of such plan, as well as longer term goals,
- (3) **make recommendations to enhance standardization and integration of data systems and data management practices** across all executive branch agencies,
- (4) 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
- (5) **set goals for improving the online repository** established pursuant to subsection (i) of this section.

Each state data plan shall provide for a procedure for each agency head to report to the Chief Data Officer regarding the agency's progress toward achieving the plan's goals. Such plan may make recommendations concerning data management for the legislative or judicial branch agencies, but such recommendations shall not be binding on such agencies.”

2021 – 2022 state data plan: focal points

The plan includes three focal points – with implementation steps and goals associated with each.

COVID-19 response
and recovery

Equity in the data
lifecycle

Using data for
decision-making

2021 – 2022 state data plan: goals and metrics

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

- *Primary metrics:*
 - Site traffic
 - Frequency and consistency of updates and metadata, particularly for high-value / high-priority data

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

- *Primary metrics:*
 - Consistent standards for data collection
 - Increased disaggregation of data
 - Improved metadata
 - Increased outreach and engagement with residents and nonprofits

Improve use of data to inform decision-making

- *Primary metrics:*
 - Reduced time to complete data sharing agreements
 - Improved customer experience for data requestors
 - Number of research partnerships
 - Examples of use of data to inform policy and practice

2021 – 2022 state data plan: implementation steps

COVID-19 response and recovery

- Prioritize access to high-value datasets on open data portal
- Improve the user experience with better navigation, user-friendly tools (like data stories), more consistent documentation
- Increase use of automated feeds to the portal and reliable update schedules
- Coordinate with planning efforts for health information technology and state information technology infrastructure
- Improve access to, coordination and management of the geospatial data for response to future emergencies

Equity in the data lifecycle

- Improve standards and documentation for gender identity, sex, race / ethnicity and other dimensions
- Develop user-friendly processes to request and access state data
- Increase reporting and disclosure of data by factors such as race, ethnicity and gender, and their intersections, within the appropriate guidelines
- Increase the formal and informal avenues for resident involvement in the data lifecycle
- Improve access to, coordination and management of the geospatial data for understanding disparities by place

Using data for decision-making

- Implement uniform interagency data sharing protocol, based in 2Gen efforts
- Use Data Sharing Playbook to support process improvements
- Build on use cases, such as Governor's Workforce Council and 2Gen
- Pursue skill and capacity-building opportunities
- Improve documentation to facilitate research partnerships
- Continue expanding P20 WIN system
- Continue supporting efforts to improve health data infrastructure
- Continue supporting efforts to improve service delivery

Agency Requirements

Agency progress and data governance elements

The chart below summarizes the steps taken by agencies to comply with [C.G.S. 4-67p](#) and to make progress on the goals of the State Data Plan.

	DAS	DCF	DCP	DDS	DECD	DEEP	DESPP	DMHAS	DMV	DoAG	DOB	DOC	DOH	DOI	DOL	DOT	DPH	DRS	DSS	DVA	OEC	OPM	SDE	*OHS	*OSC	*CTGB
Designate Agency Data Officer																										
Inventory high value data (2018)																										
Update high-value data inventory (2019)																										
Develop open data access plan																										
Document data request response process																										
Update high value data inventory (2020)																										
Update high value data inventory (2021)																										

*Voluntary compliance

High value data inventory

Agency Input Sheet (Non-GIS) High Value Data Inventory - Agency Input Version V2.0						
Agency	Inventory ID	Application/Dataset/ Data Repository	Application/Dataset Description	Business Function/Area	Dataset Structure	Major Data Elements
OPM	OPM002	Small Town Economic Assistance Program (STEAP)	The Small Town Economic Assistance Program (CGS Section 4-66g) funds economic development, community conservation and quality-of-life capital projects for localities that are ineligible to receive Urban Action (CGS Section 4-66c) bonds. This program is managed by the Office of Policy and Management, and the grants are administered by various state agencies.	Office of the Secretary/Division of Administration	Structured (Machine Readable)	Town Name, Award Amount, Award Date, Award Agency
OPM	OPM003	State Elections Enforcement Commission (SEEC) - Contracts Outside of Core	This is a list of contracts that have been executed but have not been entered into CoreCT as of the monthly report run-time. The report is in Excel format. Subdivisions of the State (such as municipalities) are not included.	Division of Administration	Structured (Machine Readable)	FEIN, Vendor Name, Contract Name and start and end date
OPM	OPM004	State Elections Enforcement Commission (SEEC) - Exclusions	This is a list of contracts that should be excluded from SEEC reporting. Those contracts that should be excluded are: • all OPM contracts with other state agencies; • all contracts with municipalities; • all contracts with other political subdivisions of the state; and • contracts that are exclusively 100% federally funded.	Division of Administration	Structured (Machine Readable)	FEIN, Vendor Name, Contract Name, and start and end date
OPM	OPM005	State Elections Enforcement Commission (SEEC) - Prospective State Contractors	A prospective state contractor is a person, business entity or non-profit organization that has submitted a response to a contract solicitation by OPM (e.g., RFP/RFQ). Prospective State Contractor information is added as a result of a contractor's submission of a response to an RFP/RFQ, but prior to being selected for or rejected for a contract/award. Prospective State Contractor information is removed when a decision is made to either award or not award a contract to a contractor on the list. The report is in Excel format.	Division of Administration	Structured (Machine Readable)	FEIN, Vendor Name, Vendor Address, Contact Name, Contact Title, Contract Solicitation Number, and start and end date

The data collection form for the high value data inventory (left) has remained largely the same since it was created in 2018.

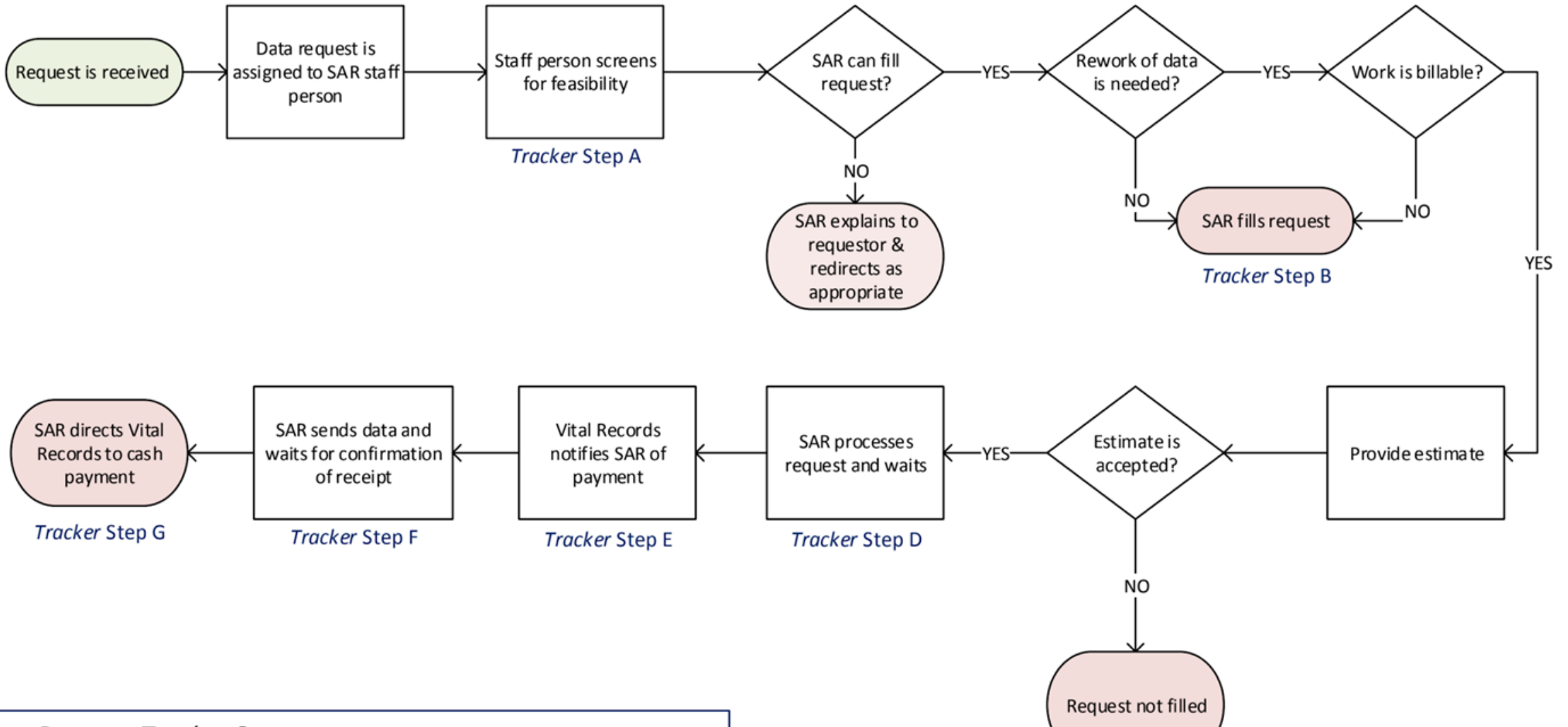
The [2020 CT Data Catalog](#) on the Open Data Portal displays GIS and non-GIS data together, whereas previous years had separate datasets for spatial and non-spatial data.

Are there changes you would like to see to the high value data inventory collection form? To the corresponding datasets on the portal?

Open data access plans and data request process documentation

- Open data access plans
 - The open data access plan form (similar to the high value data inventory form) is available on the Agency Guidance page of the State Data Plan website [here](#).
 - The purpose of the plan is to document what data is being made public by agencies and where it is published, as well as to identify data that could be made public in the future.
 - Additional guidance from OPM about data publication on the portal is being developed, and the open data access plan form may be updated as a result.
- Data request processes
 - Each agency should have documentation that explains the process for responding to data requests made to the agency.
 - Documenting these processes can facilitate easier data request processes.
- Thoughts about how to improve these resources? Or need support? Let us know!

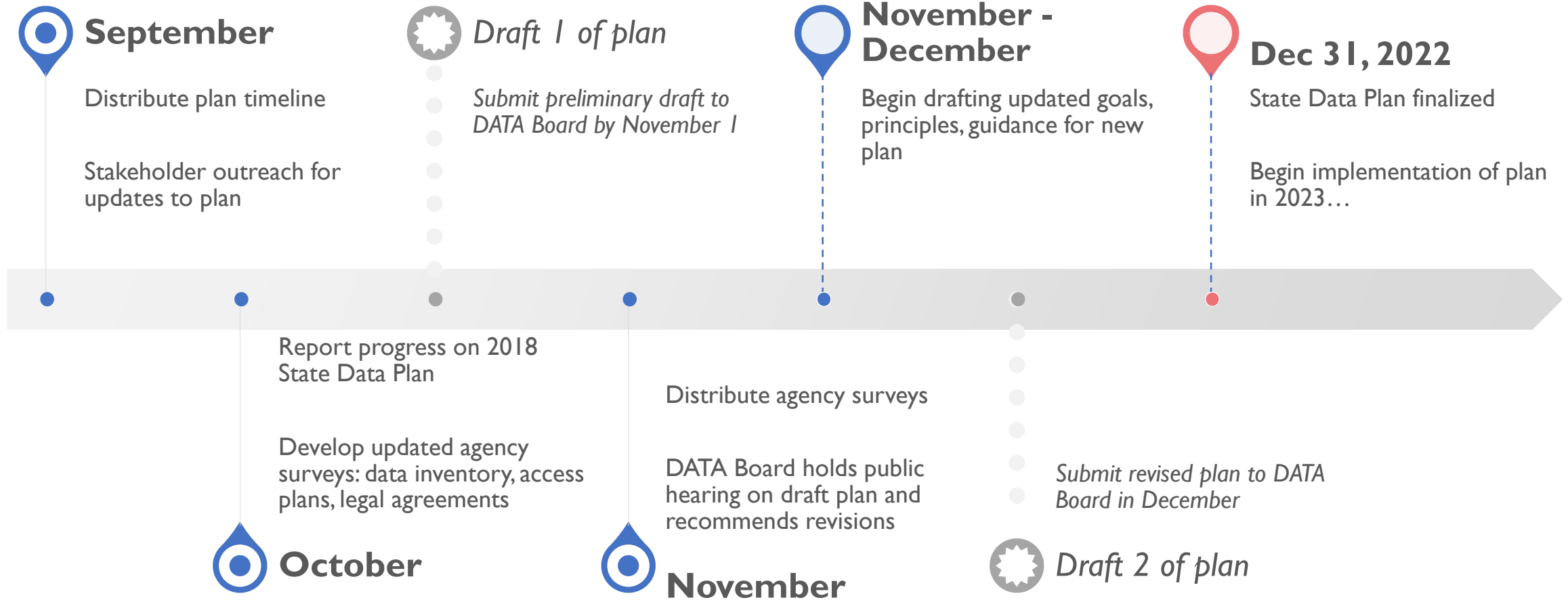
Sample Data Request Workflow



Areas for future agency focus or support

- Data publication workflow (review and approval process)
- Automation (working on documentation for tools now)
- Aggregation (cell suppression)
- Data visualization / style guide ([OEC example posted](#))
- Disaggregation by race / ethnicity / language (standards, tools)
- Deep dive or 'cohort' training
- Internal or 'closed' data portal training
- GIS clearinghouse

State Data Plan Update – Draft Milestones





Adjourn

Thank you!

Scott Gaul

Chief Data Officer

CT Office of Policy and Management

Scott.Gaul@ct.gov