

California's Data Strategy

Equipping California to navigate the data landscape

Joy Bonaguro | Chief Data Officer | State of California | August 2020

[Executive summary](#)

[Strategic plan](#)

[INTRODUCTION](#)

[Why data and this strategy matters](#)

[Mission and vision](#)

[SUMMARY OF ENVIRONMENTAL SCAN](#)

[Approach and limits](#)

[Barriers to data use and why we need to solve them](#)

[Departments: Different flavors imply different data needs](#)

[Moderate confusion on the role of business versus technology](#)

[Practices within the state](#)

[Lessons from other jurisdictions](#)

[GOALS AND STRATEGIC OBJECTIVES](#)

[Goals](#)

[Strategic objectives](#)

[Crosswalk between goals and objectives](#)

[Implementation timing and approach](#)

[CONCLUSION](#)

[Appendices](#)

[A. Acknowledgements](#)

[B. Documents reviewed](#)

Executive summary

MISSION AND VISION

As a society, we have tasked our governments with some of our most complex challenges, including educating our children, balancing public safety with social justice, and providing services to our most vulnerable. Good use of data is the tool we can use to navigate that complexity and ensure that our programs and services are working well for all Californians.

The mission of the Office of the Chief Data Officer is to empower use of data by ensuring the state has the infrastructure, processes, and people to manage, access, and use data efficiently, effectively, securely, and responsibly. By acting on our mission, we will realize our vision of better decisions, services, and outcomes for Californians through better use of data.

GOALS AND STRATEGIC OBJECTIVES

Our goals and supporting strategic objectives are structured around the analogy that in order to successfully navigate the “data landscape” we need to intentionally build the roads, craft the rules of the road, and boost the drivers. Much like in the real world, we want to avoid data roads that lead to nowhere, are poorly maintained, or confuse our drivers. Our virtual data world requires the planning and care that we put into the roads and bridges of the real world.



Goal 1. Streamline data access | Build the data roads

Objective 1. Enduring longitudinal datasets. Help accelerate and align the creation of enduring longitudinal datasets across the state. Longitudinal datasets, which track people across programs and over time, are key to understanding the effectiveness of a range of state programs - from early childcare, to education, to social services. They allow us to start to understand the relationship between services provided and subsequent outcomes related to health, education, economics, and more.

Objective 2. Open data. Assess statewide open data efforts and develop a plan to strengthen. Open data is a key piece of data infrastructure and provides a means to streamline the discovery of and access to public datasets - for both the public and state employees.



Goal 2. Improve data management and governance | Craft the rules of the road

Objective 3. Authoritative data management. Develop an approach to prioritize and support the identification, documentation, development, and distribution of authoritative

datasets. Authoritative data reduces duplication and confusion in terms of developing or procuring, discovering, accessing, integrating, and using data across the state.

Objective 4. Interagency data exchange. Implementation of the statewide data exchange agreement. Holistic, whole-person care for efforts such as homelessness or Cradle to Career relies on the secure, legal, and appropriate exchange of confidential information across departments and agencies. An umbrella data sharing agreement harmonizes security and privacy rules and reduces administrative effort for data exchange.

Objective 5. Ethical data governance and management. Develop, adopt, or modify playbooks for ethical data governance and management throughout the data lifecycle. Data must be managed throughout its lifecycle in a way that not only promotes quality, consistency, usability, and reuse but also the ethical use of data, including avoiding unintended bias or inequity. Just because you *can* use data does not necessarily mean you *should*.



Goal 3. Spur data use and ability | Boost the drivers

Objective 6. Data skills. Assess the need for data skill development and existing training programs to develop a statewide approach for data skills if appropriate. Common data skills could lift data competency across data staff, frontline employees, and leadership.

Objective 7. Data jobs. Revisit data related job classifications to ensure the state is positioned to both promote and hire modern data skills. The nature and structure of data related functions continues to evolve in terms of skills and competencies, technical expertise, and salary. We must maximize our ability to not only recruit and hire the data talent we need but provide meaningful advancement opportunities for our existing staff.

Objective 8. Data science and advanced analytics. Explore pilot centers to demonstrate the power of data science and analytics. The best way to drive use of data is to use data to help solve problems that people care about. This objective is about cultivating the demand for advanced analytics and demonstrating its power to improve state services.

Our final two objectives help support all three goals.

Objective 9. Data communities. Foster data communities and networks across the state. In order to optimize our use of data, we must work across boundaries. Both formal and informal networks can help foster these connections and unlock new data opportunities.

Objective 10. Data and evidence workgroup. Establish a working group to develop a set of recommendations to help the state accelerate use of evidence across all of its practices. Similar to the federal commission under former President Obama, the state could benefit from a wholesale review of its existing practices, policy levers, and workforce to assess where and when we can improve use of data and evidence.

Strategic plan

INTRODUCTION

Why data and this strategy matters

As a society, we have tasked our governments with some of our most complex challenges. Just a few of these include educating our children, balancing public safety with social justice, and providing services to our most vulnerable, such as those experiencing homelessness. This means that **we cannot afford to guess how well our services are working**. The lives of Californians depend on us knowing what works and what doesn't.

That's why we need robust data and evidence-based approaches to ensure that our programs and services are working in the way that we expect them to work. Better use of data and evidence by state employees and leadership can improve our decisions, services, and ultimately the outcomes and lives of our residents.

Mission and vision

The mission of the Office of the CDO is to empower use of data by ensuring the state has the infrastructure, processes, and people to manage, access, and use data efficiently, effectively, securely, and responsibly.

By acting on our mission, we will realize our vision of better decisions, services, and outcomes for Californians through better use of data.

SUMMARY OF ENVIRONMENTAL SCAN

Approach and limits

To ensure the state's data strategy was grounded in the needs of the state as well as lessons from other jurisdictions and sectors, I conducted an environmental scan to learn about the following areas:

- **The state of the state.** What are the bright spots, common challenges, and lessons learned? What existing data groups are there and how do they operate?
- **Other states and federal equivalents.** What are leading practices? Common structures? How should I think about the difference between the role of data at the state versus local versus federal level?
- **Regional / county / local.** What major initiatives should the state understand? What can the state do or not do to support sub-state work?

- **Think tanks and research centers.** What guidance is available for data strategies from both the public, university, and private sectors?

The methods I used were a mix of interviews, focus groups, data collection, and reviewing existing documents, studies, articles, and white papers.

In the early days of my plan, the crisis and response to covid became increasingly urgent. While I did still learn a great deal, including from participating in covid response, my investigation was necessarily truncated. As a result, the results below are incomplete and part of the plan for the next year is to continue a low-level environmental scan and bake research into a subset of the strategic objectives.

Barriers to data use and why we need to solve them

During the listening tour, I spoke not only with individuals but created a series of employee listening sessions, including program managers and analysts. During these sessions, I asked a series of questions about data use, including barriers. While this survey is not representative of all employees, it gives us an indication of the nature of data barriers. Moreover, the results are consistent with my experience speaking with and working in other jurisdictions. The table below summarizes what I learned with respect to the major shared barriers and the potential impacts on the state.

Shared barriers	Why this barrier is a problem
<p>Finding and accessing data across departments is the top barrier. The vast majority (88%) of state employees I spoke with noted that access to cross department data was a medium or major barrier to data use. Another 57% report that accessing data within even their department is a major or medium barrier.</p>	<ul style="list-style-type: none"> ● Limits data use. If you don't know about or can't access data, you can't use it by definition. ● Wasted time. Staff wastes time hunting down data or conversely figuring out how to supply it to data requestors. ● Potential for duplication. Staff creates new data resources that may already reside in another department.
<p>Other top barriers include data consistency and quality. Employees classified data consistency and quality as the next biggest barriers with 82% and 74% citing them as a major or medium barrier respectively.</p>	<ul style="list-style-type: none"> ● Limits combining data from different systems. Inconsistent data is hard to combine and our hardest questions require that we combine data from across different sources. ● Wasted time. Staff wastes time figuring out how to link and match inconsistent data and track down definitions or standards. ● Unnecessary effort. A lack of shared definitions or standards results in staff recreating the wheel for new data systems.

State employees cited several other barriers shared by a majority or sizable minority. Each of these require more exploration to better understand the nature of the barrier and related impacts as well as what is the best approach to addressing the barrier:

- **Staffing and staff skills are common and related barriers.** A majority of employees (57%) reported that staffing and roles for data were a major or medium barrier, while a strong minority reported that skills in using data were (43%). Both issues are worth understanding in greater detail. Do we not have the right roles? Are they structured correctly? Where do we have skill gaps?
- **Leadership support may be a common challenge.** Employees report that executive focus and support (46%) and access to decision makers (41%) are medium to major barriers. While these percentages are lower, they approach 50% and are important as the role of leadership in driving use of data and adoption of a data culture is not only critical but arguably the most important component. Lack of leadership support could result in under-resourcing the data function, delegating data responsibilities inappropriately, or reducing incentives to use data if not requested by decision makers. Understanding this challenge is critical as it could mean a range of things including that not all leaders fully understand how data can help them, challenges translating data insights into something actionable, and/or other factors.
- **Resistance to data sharing and privacy and legal concerns.** The majority found resistance to data sharing (60%) and privacy and legal concerns (56%) to also be a challenge. In practice, I've found these issues to be linked and benefit from a variety of strategies not least of which is a clear legal framework, consistent data sharing standards and processes, and ongoing education to support both. Much work has already been done in the state on these issues and our strategy below incorporates it.

Departments: Different flavors imply different data needs

In the course of learning about the state, I started to observe that there were key differences amongst departments. If we think about departments, they come in at least three different flavors or types (and sometimes a blend of these):

1. **Researcher regulator.** Primary function is to issue policies and regulations based on evidence and research and may include an enforcement arm, e.g. CalEPA departments.
2. **Funder.** Primary function is to administer, distribute, and allocate funds to regional, county, or local entities who then provide the services, e.g. California Department of Health Care Services.
3. **Service provider.** Primary function is to deliver services internally or to the public, e.g. CalHR or Department of Motor Vehicles.

While this concept is imperfect and the categories are fluid, it does help provide a construct for thinking through data services and strategies and how those may differ by agency and/or department with respect to staffing, data sources, and data use:

Department type	Existing data staff	Data sources	Data use
Researcher regulator	PhDs and statisticians; though may lack	Large complex datasets from studies or remote	Focus on studies and analyses

	dedicated data roles	sensing	
Funder	Policy and planning staff	Range of data sizes often from local entities	Focus on evaluation and assessment
Service provider	Operational staff, fewer data focused roles	Transaction systems; enterprise systems	Focus on performance management and operational analysis

These differences in turn implicate the nature and structure of the decision-making process, including users and recipients of data related work and the nature of the support needed from a statewide strategy and entity. These differences are key to adopting a statewide strategy that meets the needs of different department flavors.

Moderate confusion on the role of business versus technology

One of the other areas I uncovered was some confusion about roles and responsibilities around data management and use. Several people I spoke with and one of the topics during the listening sessions focused on which types of staff should drive data work. In particular, the role of “business” versus “technology” generates the most confusion — though this wasn’t universal. Business is an umbrella term that includes operations management, policy and planning, and program management. An analogy I use to help clarify this confusion is that of the data gym:

- **Business is the person who wants to get in shape.** They have to define their fitness goals with respect to use of data, how quickly they want to achieve them, when they want to work out, and where they want to focus. You can start at any level of fitness but only you can choose to show up, work out, and decide how much to push yourself.
- **Data teams are the trainers.** They help you understand how to use the equipment to best meet your goals. They identify opportunities to optimize and advise on how to best target problem spots.
- **Technology provides the gym and equipment.** They can house the data infrastructure and build and maintain it in response to the demand and requirements from trainers and customers as well as their knowledge and expertise on managing gyms.

The Data Gym: The roles of business, data, and technology



Business needs to decide how much effort they want to put into getting their data game into shape.

Data teams train business on how to best use data equipment and how to target trouble spots.

Technology provides and maintains the data gym.

In addition to the gym analogy, I like to ask business leaders, would you let your technology team...

- Define your program's goals and objectives?
- Define your key metrics and performance measures?
- Oversee the evaluation of your programs?
- Oversee the effectiveness of your internal operations and processes?

If they answer no, then this helps clarify that data is the responsibility of the business side. Data is the signal the business needs to improve programs and manage work.

In short, management and use of data cannot be abdicated to technology staff or data teams. We will weave these messages into our various objectives both in communications, policy, and practice.

Practices within the state

While my review of state practices was hampered by Covid-19 response, I was able to speak to a variety of individuals, listen in on many meetings, and review various data related documents. While my education is far from over, I was able to learn of many exciting efforts across the state, including but not limited to the following:

- Multiple data governance initiatives, including Caltrans, Franchise Tax Board, the AB 1755 Partner Agency Team focused on water data, Health and Human Services Agency, and more with several spanning multiple departments or even agencies.
- A variety of data literacy initiatives including both assessment of staff data literacy, training efforts, and mentoring programs.
- Communities of practice both within and cross agency, whether focused on open data, GIS, or other topic areas.
- Emerging data leadership at the agency and department level, including the creation of a CDO in both the Health and Human Services Agency and the Department of Motor Vehicles and the Data Strategy role in the Department of Social Services.
- The expansion of open data efforts, with the passing of the open data policy last spring and the creation of the GIS portal in late fall.
- Comprehensive dashboarding efforts in the Department of General Services and the State Water Resources Control Board to name just a few.
- The scaling of an umbrella cross agency data sharing agreement to harmonize requirements and streamline administration.
- The strong push to create integrated data systems in multiple content areas.

Collectively, these efforts form a strong foundation for a statewide strategy to leverage in terms of lessons learned, scaling, and implementation.

Lessons from other jurisdictions

As part of my scan, I reviewed practices and approaches both across the US at the state and federal level as well as internationally, including Canada, New Zealand, the UK, and Australia. Appendix B lists many of the documents reviewed. In some cases I spoke with individual representatives from the jurisdiction or individuals who were knowledgeable across multiple jurisdictions, including colleagues via the [Beeck Center's State Chief Data Officer Network](#). My high-level lessons include:

- Having a data focused strategy is emerging as a common practice, though strategies varied in both their detail and focus.
- Several jurisdictions use legislation to specify data efforts to some degree, including requiring a report or strategy.
- Most national efforts are in their early stages and as a result, did not include detailed information on prior efforts and results.

We will continue to participate in state level networks and monitor national efforts across the world to ensure that we learn from and leverage the lessons of our peers.

GOALS AND STRATEGIC OBJECTIVES

Goals

Based on the environmental scan and past experience, the following high-level goals will help advance the data game across the state. These goals may evolve over time but will likely stay with us for the next several years. We expect these goals to lead to a handful of key results, including both the supply of and demand for high quality data. Ultimately, this will help us improve decisions, services, and outcomes for all Californians.

High level goals	Key results	Ultimate results
<ol style="list-style-type: none"> 1. Streamline data access 2. Improve data management and governance 3. Spur data use and ability 	<ol style="list-style-type: none"> 1. Data when and where you need it 2. Secure and appropriate use of data 3. Higher quality data that is more consistent 4. Increased demand for data <i>and</i> the ability to meet that demand 	<ol style="list-style-type: none"> 1. Better decisions 2. Better services 3. Better outcomes

Our goals and supporting strategic objectives are structured around the analogy that in order to successfully navigate the “data landscape” we need to intentionally build the roads, craft the rules of the road, and boost the drivers. Much like in the real world, we want to avoid data roads that lead to nowhere, are poorly maintained, or confuse our drivers. **Our virtual data world requires the planning and care that we put into the roads and bridges of the real world.** If not, we will continue to ask both our leaders and our employees to navigate the data landscape without a road, a map, or even a compass.

Below I describe each goal in greater detail.

Strategic Goals: Equipping ourselves to navigate the data landscape



Build the data roads
streamline data access



Craft the rules of the road
improve data management and governance



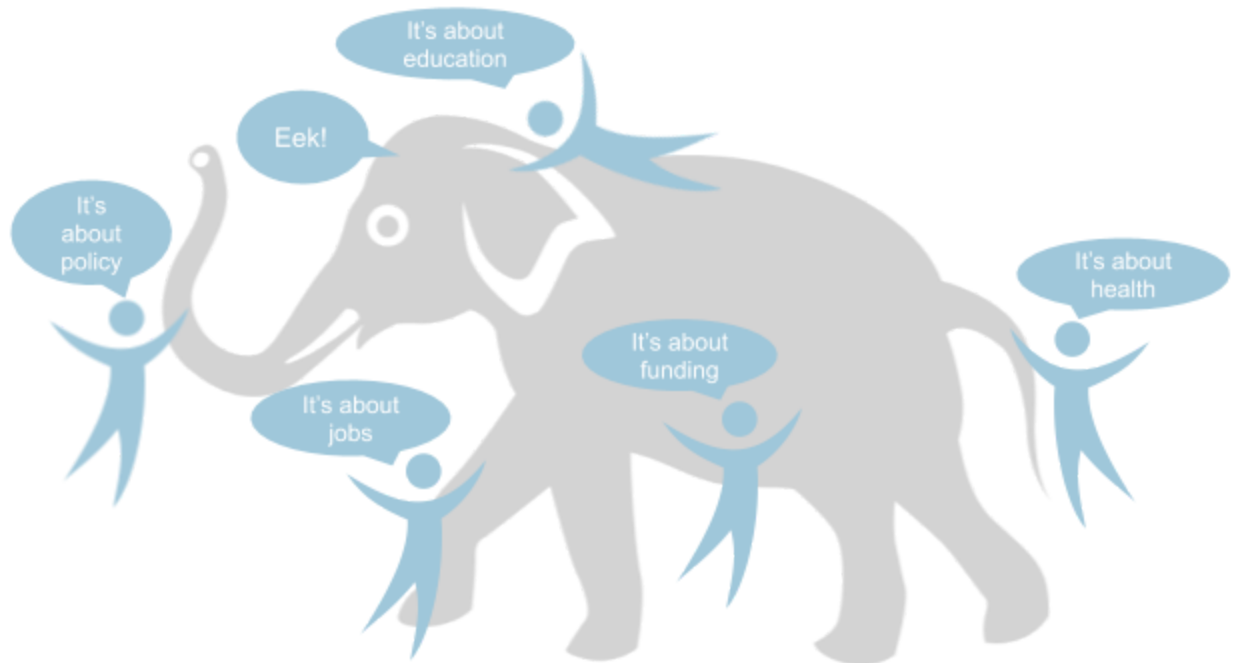
Boost the drivers
spur data use and ability

Goal 1. Streamline data access | Build the data roads

As mentioned in the environmental scan section, access to data remains a challenge both within and between departments. Like the proverbial blind men and the elephant - if you only see your data piece, you can't possibly understand the full challenge or more importantly, the full solution. Access to comprehensive, integrated data is about understanding the whole elephant, not just your piece. For example, for people experiencing homelessness, they may interact with multiple systems beyond homeless services including health care, social services, law enforcement, and emergency response. The challenges of our most vulnerable do not respect bureaucratic boundaries. Every time we focus on just one piece, our ability to deliver the best services we can to Californians suffers.

Access to data both within and across departments is difficult due to technology barriers, legacy systems, and lack of shared data infrastructure. Our strategic objectives focus on reducing these barriers and moving towards shared, well governed data infrastructure.

Why data access matters: If you only see part of the problem, you may solve for the wrong thing



Goal 2. Improve data management and governance | Craft the rules of the road

Data management and governance are big and often boring terms. But they are essential. The evolution of stop signs provides an apt and relevant example. Prior to the 1968 Vienna convention, stop signs varied tremendously across the world. Even adopting this convention took - in huge part - the prompting of the adoption of a standard in the US in 1966 *after decades of disagreement*. We don't even question the wisdom of why we need standards in road signs

and usage. Not only is a matter of safety, it is also a matter of efficiency and convenience. We can easily navigate unknown territories and drive successfully without in depth training.

However, when using datasets across systems, we encounter a new, non-standard world at every turn. We can extend the analogy even further - can you imagine a US highway system where the highways ended at each border? Of course not. It would be a huge inefficient waste. But that's exactly what we experience in the data realm. Each system operates as its own state with its own roads that lead to dead ends at each border. Unfortunately for Californians, their needs are international so to speak.

And like actual roads, we need rules about safety and security. While standard, integrated, comprehensive data is needed to better understand our challenges, it must be done in a way that is secure, legal, and appropriately respectful of the data the public has entrusted to us. We need the right onramps, seatbelts, and recalls when using sensitive data.

We've neglected the standards of the digital world. And while the impacts are less obvious they are no less important.



Goal 3. Spur data use and ability | Boost the drivers

Data is a team sport. Data literacy and the skills and capacity to use data are needed for frontline employees as well as executive leadership. While our survey suggested that data skills are not the biggest blocker to data use, this goal speaks to ensuring not only the right data focused roles and hands-on skills but also more broadly to the organizational roles and structure that will encourage broader and deeper data use.

For example, Canada's federal data strategy is investigating how to formalize and standardize the use of data throughout key decision-making processes, including budget requests and official memos. Other countries, such as Australia, are developing robust standards for program and policy evaluation as well as revisiting the need for laws that specifically target use of data

and data sharing. The very fact that these conversations are happening across several countries, including at our own federal level with the release of a national data strategy, speaks to the relative immaturity of the data practice in contrast to financial, capital, and workforce planning.

Strategic objectives

Below is a set of strategic objectives designed to support our goals. We expect to prioritize these per the implementation approach discussed in the next section. Some of these objectives require additional research before determining the best path forward. When appropriate, we list key partners needed to make the objective successful.

Objective 1. Enduring longitudinal datasets. Help accelerate and align the creation of enduring longitudinal datasets across the state.

Longitudinal datasets, which track people across programs and over time, are key to understanding the effectiveness of a range of state programs - from early childcare, to education, to social services. They allow us to start to understand the relationship between services provided and subsequent outcomes related to health, education, economics, and more.

While many states across the country have deployed longitudinal data systems, California is starting on this journey and is in the midst of several initiatives focused on creating persistent and durable datasets, including the Homeless Data Information System, the Cradle-to-Career planning process, and the research data hub in CHHS. It is a unique time and opportunity to ensure these nascent investments build upon and leverage one another, especially given recent budget constraints as well as the potential disproportionate impact of covid on our most vulnerable populations.

Key partners. *Cradle-to-Career partner entities, California Business, Consumer Services and Housing Agency, California Health and Human Services Agency.*

Objective 2. Open data. Assess statewide open data efforts and develop a plan to strengthen.

Open data is a necessary piece of data infrastructure and provides a means to streamline the discovery of and access to public datasets - for both the public and state employees. For several years, a variety of departments have made tremendous progress in open data with more than ten portals and hundreds of datasets available. And in March of 2019, the State passed a statewide open data policy and supporting handbook to help unify efforts across the state. With more than a year since the policy became active, we have an opportunity to assess what's working well, what's not, and what is the next stage in terms of continuing to improve and evolve our approach to open data. A significant part of this will be engaging our public users of data and ensuring their input and voice helps inform our plan.

Key partners. Department and agency open data programs and governance entities, CDT's open data portal team, existing and potential users of open data.

Objective 3. Authoritative data management. Develop an approach to prioritize and support the identification, documentation, development, and distribution of authoritative datasets.

Authoritative datasets (i.e. canonical datasets such as addresses, department names, as well as data standards) must be readily available and well understood in the state. A shared understanding of what data or standard is authoritative is key to reducing duplication and confusion in terms of developing or procuring, discovering, accessing, integrating, and using data across the state. Moving towards authoritative data includes identifying the official source (or developing it if not available), ensuring that data is well understood and documented, and then finally, providing that data as a service so it can be easily consumed across the state (i.e. making it available on demand to both people and computers). This effort will require the input of various groups across the state and is not a single year objective but an enduring programmatic effort.

Key partners. Department governance programs, GIO, CalData, GIS Working Group.

Objective 4. Interagency data exchange. Implement the statewide data exchange agreement.

Social service delivery is in the midst of a migration from program to people centric whole person care. Our most vulnerable individuals touch multiple systems - education, human services, and criminal justice - which have historically operated in silos. The transition to coordinated care will better meet the needs of our residents by tailoring care to meet the needs of each individual, rather than administering programs with a one-size-fits-all approach and requiring individuals to navigate complex government silos. However, this relies on the secure, legal, and appropriate exchange of confidential information across departments and agencies. Historically, negotiating these agreements has resulted in lengthy discussions and extensive administrative overhead. The Interagency Data Exchange Agreement creates an umbrella agreement for data exchange that not only ensures that each data exchange is legal, appropriate, and secure but also streamlined and efficient.

Key partners. All Agencies and data governance programs, CDT.

Objective 5. Ethical data governance and management. Develop, adopt, or modify playbooks for ethical data governance and management throughout the data lifecycle.

Data must be managed throughout its lifecycle in a way that not only promotes quality, consistency, usability, and reuse but also the ethical use of data. Just because you *can* use data does not necessarily mean you *should* use data. Moreover, as the state continues to adopt advanced statistical methods, e.g. machine learning, we need to ensure that our use of these tools is appropriate and ethical and helps reduce bias and inequity. A set of shared playbooks can both promote standard practices and ease the path to adopt what are often poorly defined concepts. A variety of playbooks already exist both internally and externally that we can use or

adapt, and we may identify the need for new ones. These playbooks will complement and enhance our other strategic objectives, in particular those of authoritative data management, open data, data exchange, and longitudinal datasets.

Key partners. Department governance programs, CalData, GIS Working Group.

Objective 6. Data skills. Assess the need for data skill development and existing training programs to develop a statewide approach for data skills.

A variety of training opportunities exist in the state, including classes from CalHR and California Department of Technology. In addition, several departments have developed their own data themed curriculums or have assessed the need for data skills. We will review these existing efforts as well as assess potential demand for additional training or skill development to develop recommendations for how to proceed on training and development.

Key partners. Department training programs, CalHR.

Objective 7. Data jobs. Revisit data related job classifications to ensure the state is positioned to both promote and hire modern data skills.

The state uses a variety of data-related roles, including the research specialist and analyst roles. Prior efforts simplified the specialist and analyst roles into a single job classification. However, the nature and structure of data related functions has evolved even in the last few years both in terms of skills and competencies required, level of technical expertise, and competitiveness of salary ranges. A review of our job classifications will help inform how the state should continue to evolve its job classes to maximize our ability to not only recruit and hire the data talent we need but provide meaningful advancement opportunities for our existing staff.

Key partners. CalHR with input from Departments and data teams.

Objective 8. Data science and advanced analytics. Explore pilot centers to demonstrate the power of data science and analytics.

The best way to drive use of data is to use data to help solve problems that people care about. Data science uses advanced statistical methods, including machine learning, to bring insights and new tools to those problems. With earlier work under a program called [DataScienceSF](#), we developed a [methodology](#) to consistently solicit and deliver on data science projects to solve operational problems. While many of our objectives are focused on creating the supply of data - this objective is about cultivating the demand for data.

Specifically, we will explore a pilot center focused on operational analytics. An operational analytics team could focus on soliciting, scoping, and executing on data science projects in departments that provide services. The goal would be to help them streamline and improve service delivery and to do so in a way that is fair and equitable. Unlike the [funder and researcher regulator departments](#), service providers are less likely to have staff dedicated to

data science, increasing the potential value of the pilots. This team could also help support and evaluate other initiatives related to government transformation. We will explore how this could be created, where it could sit, and how it could determine the projects it pursues.

Objective 9. Data communities. Foster data communities and networks across the state.

People networks are the threads that can stitch together cross departmental and cross agency work. We all need each other to optimize our use of data and no entity is an island. Both formal and informal networks can help foster these connections and unlock new data opportunities. There are a handful of existing data communities, including CalData and the GIS working group. We need to continue to foster these as well as develop additional groups and mechanisms for engaging, especially given the sharp increase in telecommuting during covid.

Key partners. Department governance programs, GIO, CalData, GIS Working Group.

Objective 10. Data and evidence workgroup. Establish a working group to develop a set of recommendations to help the state accelerate use of evidence across all of its practices.

While each of our strategic objectives is important, the state would benefit from a wholesale review of its existing practices to assess where and when we can improve use of data and evidence. The Federal Government commissioned a similar assessment under President Obama that eventually resulted in a variety of changes, including creating the role of the Chief Data Officer across entities and developing robust approaches to evaluation.¹ Moreover, my review of state programs suggests a range of models we could assess and adopt in part.

However, it would be both inappropriate and inefficient to do this alone and it would waste an opportunity for shared learning. A California wide group could leverage the expertise of state employees, local and county representatives, academics, and other sectors to learn from efforts at the federal and state level as well as other countries. This group would be tasked with developing recommendations, including the best way to implement them, on topics such as:

- How can the State best enhance the use of data and evidence in existing policy and funding decisions?
- What staffing roles and structures are most effective in terms of increasing effective use of data in departments and agencies?
- What type of additional shared data approaches, if any, are needed to improve cross agency and cross department work?
- To what extent should the state standardize, scale, and adopt practices such as robust evaluation, randomized controlled trials, and performance management?
- How can the state best foster research partnerships that are relevant to agency and department questions and priorities?

¹ Similar efforts are underway in Canada, Australia, and New Zealand.

Crosswalk between goals and objectives

Our strategic objectives support each of our goals to some degree. However, each objective varies in how much it supports the goal. For example, while data communities will provide some level of support to each goal, data skills and jobs are focused on spurring data use and ability. The table below indicates the degree to which each objective supports each goal.

Objective	Streamline data access	Improve data management and governance	Spur data use and ability
Enduring longitudinal datasets	High	Medium	Low
Open data	High	Medium	Low
Authoritative data management	Medium	High	Low
Interagency data exchange	Medium	High	Low
Ethical data governance and management	Medium	High	Low
Data skills	Low	Medium	High
Data jobs	Low	Medium	High
Data science and advanced analytics	Low	Medium	High
Data communities	Medium	Medium	Medium
Data and evidence workgroup	TBD	TBD	TBD

Implementation timing and approach

We cannot pursue all of our objectives simultaneously. As a result, we will prioritize effort and focus due to either resource availability, dependencies on existing efforts and timing, or the need for additional planning or assessment. The proposed implementation timing and approach reflects our best planning guess due to these factors and is not a reflection of inherent priority.

Our primary focus for the next 6-12 months will be on the following objectives:

- Enduring longitudinal datasets. Planning or procurement is already underway for both the Homeless Data Information System and Cradle-to-Career system, which makes this objective a priority by default.
- Interagency data exchange. The agreement and its implementation are needed for the longitudinal data system investments and so follow that priority.
- Data communities. Fostering data communities will help inform and scope much of the rest of our work. As a result, it's an early priority to help shape the rest of the objectives.

The following objectives, we will start on but scope to a manageable effort, under partnership with others, and with the goal of either expanding over time or scoping into digestible projects:

- Open data. We would like to start with an assessment of the program and expect some effort to result from the assessment starting in the second six months of the plan.
- Authoritative data management. We already have a few projects in this area. We expect this to be a series of projects of varying size and complexity. As this will be done in partnership with data stewards, we expect it to be an ongoing low level activity.
- Ethical data governance and management. Similar to authoritative data management, we will scope this to be an ongoing activity as defined by small, scoped projects.
- Data skills and jobs. These are related and require an assessment to determine what is needed. As a result, we plan on assessing these in the second half of the year to inform projects and deliverables in the last six months of the plan.

The following objectives depend on a planning process to determine the best path forward, including viability and feasibility:

- Data science and advanced analytics. The ability to pursue this objective is dependent on a resource strategy, especially given recent budget impacts due to Covid-19, and requires planning to determine feasibility.
- Data and evidence workgroup. Our initial approach will be to plan how we would like to do this, including an environmental scan as well as obtaining disparate input into what this objective should look like in practice.

The table below summarizes our implementation approach above and the color intensity indicates relative effort we expect over time.

Objective	0-6 Months	6-12 Months	12-18 Months
Enduring longitudinal datasets	Planning / implementation		Ongoing
Open data	Assessment	Implementation / ongoing effort	
Authoritative data management	Ongoing program effort		
Interagency data exchange	Planning / implementation		Ongoing
Ethical data governance and management	Series of defined projects		
Data skills		Assessment	Implementation
Data jobs		Assessment	Implementation
Data science and advanced analytics	Planning	TBD	
Data communities	Ongoing effort		
Data and evidence commission	Planning	TBD	

CONCLUSION

Crafting the first data strategy for an entity as vast and complex as the state of California is necessarily a best guess. We will learn and iterate as we implement the strategy. However, having a strategy creates a framework for aligning data efforts around long-term goals. And alignment is key to leverage and scale both existing and future efforts to improve our use of data across the state and ultimately the lives of all Californians.

Appendices

A. Acknowledgements

First, thanks must go to the entire GovOps Agency family with a special thank you to the leadership of Secretary Richardson and Undersecretary Lee. The GovOps team is warm, welcoming, smart, and fun. Shoutouts to Krista Canellakis and Angie Quirarte for providing crucial advice, wisdom, and encouragement. Stuart Drown served as my state sherpa, providing me with guidance, context, and invaluable perspective.

The vibrant CalData and GIS community was a source of input and perspective during the listening tours and will be an invaluable sounding and learning board for the future. Also, many thanks to the multiple agency groups including those of CHHS, CalEPA, the AB 1755 Partner Agency Team and more for letting me listen in and learn about the great work and thinking across the state.

Thanks go to many individuals (my apologies for anyone I missed) for taking the time to speak with me directly and share their thoughts and perspectives: Amy Tong, Udaya Patnaik, Justin Cohan-Shapiro, Mike Wilkening, Jennifer Schwartz :-), Richard Rogers, Manveer Bola, Isaac Cabrera, Scott Gregory, Walter Yu, Chad Baker, Scott Christman, Tyler Kleykamp, Greg Gearheart, Rafa Maestu, JP Petrucione, Tamara Srzentic, Tom Lupo, Tim Garza, Eric Hysen, Joel Riphagen, Cori Zarek, Jane Wiseman, Andrew Nicklin, Sasha Wisotsky Kergan, Andy Krackov, Barbara Cohn, Michael Valle, Adam Dondro, John Ohanian, Linette Scott, Meredith Lee, Jed Herrman, Joaquin Esquivel, Betty Jablonsky, Anthony Crawford, Jason Kenney, Angela Kranz, Maksim Pecherskiy, Jason Lally, Blake Valenta, Erica Finkle, Todd Reyes, Kevin Koy, Debbie Franco, Anne Neville-Bonilla, Emily Putnam-Hornstein, Linette Scott, Irena Asmundson, Melissa Whitehouse, Brent Jamison, Andrew Sturmfels, David Harris, Mary Ann Bates, Marko Mijic, Adam Readhead, Matthew Case, Lilian Coral, Laura Meixell, John Correllus, Inger Brinck, Chris Ozuna, and Ben Flores.

Thanks to the following institutions for contributing to the collective knowledge of how to make data work in government: Results for America, The Beeck Center and the state CDO Network, Harvard Datasmart, Johns Hopkins Center for Government Excellence and others.

B. Documents reviewed

Below is a good faith, though likely incomplete, list of documents reviewed in the course of writing this document. Many thanks to the respective authors for putting their work out there so the rest of us can benefit.

- Blueprint for Delivering Results in State Government. Results for America. Available from <https://blueprint.results4america.org/>.
- 2019 Invest in What Works State Standard of Excellence. Results for America. Available from <https://2019state.results4america.org/>.

- Connecticut State Data Plan. Available from <https://portal.ct.gov/-/media/CT-Data/Connecticut-State-Data-Plan-Final-pdf.pdf>
- Oregon Draft Data Strategy. Available from <https://www.oregon.gov/das/OSCIO/Pages/DataStrategy.aspx>.
- Canada Data Strategy Roadmap. Available from https://www.canada.ca/content/dam/pco-bcp/documents/clk/Data_Strategy_Roadmap_ENG.pdf.
- North Carolina's Government Data Analytics Center (GDAC) Annual Report. Available from https://files.nc.gov/ncdit/documents/files/2017_GDAC_Legislative%20Report.pdf.
- Indiana's Management Performance Hub 2018 Annual Report. Available from <https://www.in.gov/mph/files/MPH-2018-Annual-Report.pdf>.
- Government of the District of Columbia's Chief Data Officer Annual Report 2020. Available from <https://octo.dc.gov/sites/default/files/dc/sites/octo/publication/attachments/EDI-Chief-Dat a-Officers-Annual-Report-2020.pdf>.
- Data Strategy and Roadmap For New Zealand. Available from <https://www.data.govt.nz/assets/Uploads/data-strategy-and-roadmap-dec-18.pdf>.
- The Evolving Role of the State Chief Data Officer: A Framework for Today. Beeck Center. Available from <https://beeckcenter.georgetown.edu/report/the-evolving-role-of-the-state-chief-data-office-r-a-framework-for-today/>.
- How States Use Data to Inform Decisions. The Pew Charitable Trusts. Available from <https://www.pewtrusts.org/en/research-and-analysis/reports/2018/02/how-states-use-data-to-inform-decisions>.
- The Federal Government Data Maturity Model. Available from <https://my.usgs.gov/confluence/download/attachments/624464994/Federal%20Government%20Data%20Maturity%20Model.pdf?api=v2>.
- Many documents from the Australian Government with a wealth of forward thinking. Most discoverable here: <https://www.pmc.gov.au/public-data> and here: <https://www.datacommissioner.gov.au/>.
- Big Data and AI Executive Survey 2020. Available from <http://newvantage.com/wp-content/uploads/2020/01/NewVantage-Partners-Big-Data-and-AI-Executive-Survey-2020-1.pdf>.
- Data Sharing and Analytics Governance Structure for the Commonwealth of Virginia. Available from <https://www.administration.virginia.gov/media/governorvirginiagov/secretary-of-administration/DSAAC-Data-Governance-Framework-Report-v1.pdf>
- Federal Data Strategy. Available from <https://strategy.data.gov/>.