Cambridge Open Data Strategic Plan 2026–2028

Introduction: Purpose and Context

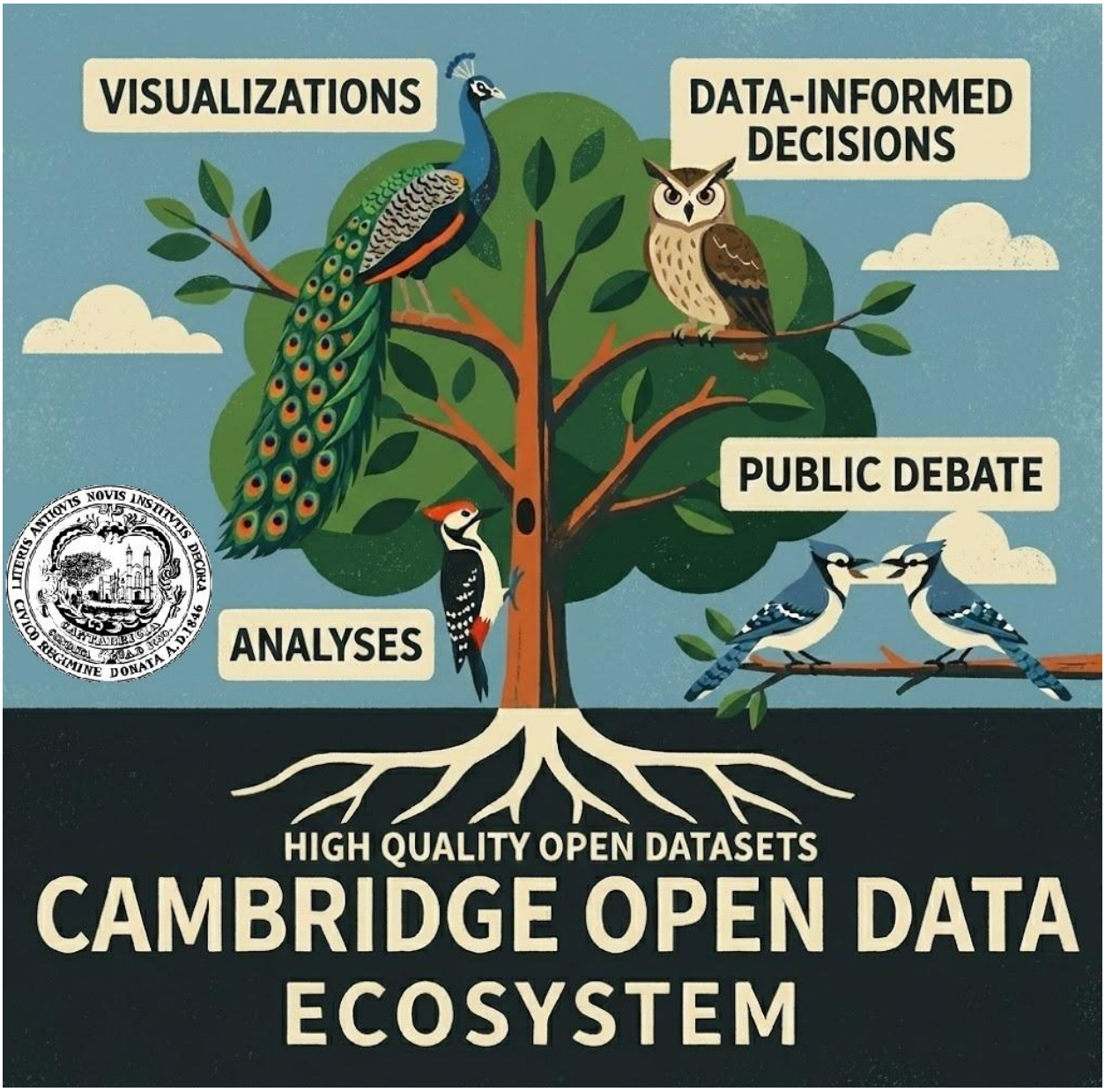
The Data Analytics & Open Data Program (hereafter referred to as *the Open Data Program*) is the City's central hub for sharing and using data to improve transparency, collaboration, and decision-making. The program manages the Cambridge Open Data Portal—the public website where anyone can access hundreds of City datasets on topics ranging from property assessments to bicycle counts. In addition to publishing data, the program supports departments across the City in analyzing information, building dashboards, and developing data-informed practices.

The Cambridge Open Data Strategic Plan for Fiscal Years 2026–2028 is a roadmap for the next three years of the City's Open Data Program. It is written for both the public and City staff, aiming to be clear and accessible. This plan builds on our previous strategic plans (2020–2022 and 2023–2025) while shifting focus to strengthen the fundamentals of Cambridge's "data ecosystem." **We use the metaphor of a healthy ecosystem to frame our approach, with a special emphasis on the "soil" that nourishes everything else.** In the context of open data, the soil represents the high-quality, sustainable datasets that Cambridge Open Data Program uniquely provides as a public good. Others (from researchers to entrepreneurs to city departments) can plant seeds and grow insights in this data ecosystem, but **only the City's Open Data Program can cultivate the rich soil of data on which all those efforts depend.** By cultivating this foundational soil, we enable the entire data ecosystem to flourish.

This plan comes at an important juncture. Cambridge's Open Data Program was one of the first of its kind when it launched over a decade ago, and over the years it has helped dozens of City departments leverage data to improve services and accountability. Our last plan (2023–2025) guided a period of growth in data analytics, governance, and community engagement, even as the COVID-19 pandemic forced adjustments. Now, as we plan for 2026–2028, we are reflecting on what has worked and what has changed, and responding to what our community needs from open data today. New technologies, especially the rise of generative AI, are reshaping how data is used and understood, adding urgency and new possibilities to our work. At the same time, the national climate for open data has become more uncertain, with increased pressure on data-gathering institutions and politicization of data sources whose mission is to provide objective information. This makes Cambridge's role as a provider of open, reliable, and accessible local data more important than ever, and reinforces the need for Cambridge to double down on this basic public good.

This plan refocuses on the basics. Dashboards and analyses, however valuable, depend entirely on having reliable, accessible underlying data. What's more, the technological landscape has shifted dramatically. Powerful visualization and analysis tools, including AI-powered capabilities that didn't even exist when our previous plans were written, are now

widely accessible to the general public. When anyone can create sophisticated charts, run complex queries, or use AI to generate insights from data, **our unique value** lies not in creating finished analyses, but in providing the high-quality datasets that make all those analyses possible. Going forward, **we will prioritize the foundational work that only we as city government can do: publishing and stewarding open datasets**. That means more data, from more departments, of higher quality, with better metadata, updated more frequently, and retired when no longer relevant. Note that this is *roots first, not roots only*. With a strong data foundation in place, both the public and City staff can innovate in countless ways. By strengthening the roots of our data ecosystem, we ensure that all the growth we hope to see (better decisions, new apps and analyses, informed public debate) can take hold and flourish.



Crucially, this plan was shaped by extensive input from the Cambridge community and experts. In mid-2025, we conducted a public survey to hear what matters most in open data. We asked residents, City employees, researchers, and others to weigh in on potential priorities, from publishing more high-quality datasets to using AI responsibly to hosting regular public workshops. The feedback was clear: all of these areas are important—plus one we hadn't explicitly asked about. Participants strongly endorsed the need for more accessible, high-value data (83% rated this "very important"), for transparent communication and engagement (75%), for helping City departments use data effectively (86%), for responsible use of new technologies like AI (52%), and for collaboration with regional and national partners (58%). But survey respondents also raised significant concerns about **privacy**. That feedback helped shape this plan's second strategic priority, which focuses squarely on data governance, privacy and public trust. In the sections that follow, we outline six strategic priority areas, each informed by what we heard and by lessons from our past work. Each priority area includes concrete, realistic goals and examples to illustrate how we will put this strategy into action.

Before proceeding**, we want to acknowledge Josh Wolff, Cambridge's first Open Data Manager, who built this program from the ground up** and authored the first two strategic plans. This plan builds on the strong foundation he established.

A note on priority ordering: These six priorities are not ranked by importance. Rather, they are organized to build logically from foundational infrastructure (data quality) through governance and engagement to internal capacity and innovation. If anything, empowering City departments (Priority 4) may be the most consequential, as internal government use of data drives better services for residents. The public's primary interest in open data is having a well-functioning, efficient, fair government—and that comes primarily through how city staff use data to improve operations and decision-making.

Strategic Priority 1: Cultivate a Sustainable Data Ecosystem (Focus on the "Soil")

Overview: The first and foremost priority is to strengthen the City's data ecosystem by building a rich and sustainable foundation of open data. In practical terms, this means refocusing on the "basics": the soil and root systems that nourish all data-driven work. The Open Data Program's unique responsibility is to collect, publish, and maintain

datasets about the city (from budget information and traffic patterns to public health metrics and beyond). This data is the fertile soil in which others can plant ideas and grow solutions. As advanced analytics tools and AI become more widespread, even non-experts can perform sophisticated analyses, **but only if the high-quality data exists and is**

accessible. Therefore, our top priority over the next three years is to ensure that Cambridge's open data is plentiful, reliable, and easy to find, so that anyone can use it to create insights and value.

Building on Past Work: In previous plans, we devoted significant effort to analytics projects and dashboards. Those efforts yielded important insights, but they also taught us a crucial lesson: **there is no substitute for making the underlying data available**. A beautiful dashboard or report might showcase a problem or solution, but if the data behind it isn't published, the wider community can't build upon that work. This plan deliberately prioritizes publishing and improving datasets **before** pursuing polished analyses. We reaffirm that open data is a core civic asset. In an era when open data cannot be taken for granted everywhere, Cambridge will reassert and expand its commitment to this fundamental public good. In short, we will concentrate on what only we can do: provide the high-quality data that fuels transparency and innovation.

Goals and Actions: Refocusing on the "roots" of our data ecosystem involves several concrete goals:

- **Expand the Open Data Portfolio:** We will publish more datasets and fill critical gaps in data coverage. Public feedback underscores the demand for more high-value datasets on pressing local issues. To identify these priorities more methodically, we will conduct gap analysis comparing our offerings to peer cities, embed feedback and suggestion mechanisms throughout the portal and outreach materials, analyze current usage patterns to understand what types of data are most valuable, and work directly with City departments and community groups to surface needs. Our aim is not just a higher quantity of data, but data that matter to residents and researchers.
- **Ensure Quality and Sustainability:** It is not enough to simply publish more data; the data must also be accurate, up-to-date, and well-documented. We will continue strengthening automation of data pipelines, data standards, and automated quality assurance measures. We will establish clearer standards and guidelines for all open datasets, covering metadata quality, formatting, update frequency, and documentation. This includes ensuring our data formats are compatible with common external tools and platforms—for example, ensuring that all datasets with geographic coordinates store them in a way that works seamlessly with mapping applications like Google My Maps. We will also work with the City's Language Access Manager to ensure accessible, plain-language descriptions (as we did for our recent community survey). We will build out automated "data QA" test suites that run on a regular schedule to check data and metadata quality—flagging issues like missing values, out-of-range entries, incomplete metadata fields, or datasets that haven't been updated on time.
- **Improve Discoverability and Metrics:** Having great data is of little use if people can't find it or don't know what exists. We will make the Open Data Portal easier to navigate and search, organizing datasets in intuitive ways and highlighting new or

popular datasets. (Our regular public communications, described in Priority 3, will also play a key role in making people aware of what's available.) We already track internal metrics on our open data program like the number of datasets, update frequency, and usage. Over the next three years, we will expand these metrics, adapt them for public audiences, and publish dashboards that show how our data ecosystem is growing (or where it needs improvement). Internally, we will continue monitoring these indicators to spot issues (e.g., a dataset that hasn't been updated on schedule) and to hold ourselves accountable for sustaining the quality of our data offerings. The metrics described at the end of this plan provide more detail on how we will measure progress.

- **Formalized "Weeding."** As our data repository grows, so too does the number of inactive or obsolete datasets. Some are of historical interest and, appropriately tagged, should be preserved. Without such careful weeding, active datasets of current interest will be drowned out and harder to find. Our data retirement guidelines were a good first step in this direction. We will work with the government open data community to adopt and participate in best practices in this area.
- **Dataset lifecycle and historical continuity.** Weeding is only the first step. As datasets age and change, we want them to remain useful to the public, especially for long-term and time series questions. For datasets that are updated over time, we will more consistently use Socrata's dataset archiving feature so prior versions can be retrieved when needed. For datasets that are no longer active but still valuable, we will keep them available with our existing **Historical** tag and with clear context about what they represent and when they were last updated. We will also treat published dataset structures as a kind of public interface: when possible, we will encourage stable field names and meanings over time, and when changes are necessary, we will aim to document them plainly so users can understand what changed and why.

Outcomes: Cambridge's approach to open data will continue to emphasize openness with responsibility. We have made hundreds of datasets public and will continue to maximize openness, including adopting open licensing to remove barriers to reuse, while using sensible safeguards for sensitive information. In 2024, for example, Cambridge formally adopted the Public Domain Dedication and License (PDDL) for appropriate datasets, reaffirming that city data should be free to reuse without legal friction. At the same time, we use techniques like data anonymization for datasets that involve personal or sensitive information. (More on privacy in Priority 2 below.) This balance of "**open by default, protected when necessary**" underpins all our data publishing work.

Strategic Priority 2: Strengthen Data Governance and Protect Privacy

Overview: As we expand and refine Cambridge's open data, we must also continue to strengthen our data governance practices to maximize data's value while minimizing risks to privacy and security. Building public trust is paramount. Residents should be confident that the City will be transparent with data, appropriately safeguarding truly sensitive

information while ensuring that information necessary for public accountability remains accessible. This balance is particularly important in government data: permit holders, property owners, and recipients of public services may need to be identifiable for the public to assess whether government is acting fairly and in the broader public interest. This priority focuses on the policies, practices, and ethical standards that ensure we share data responsibly. Good data governance means having clear rules and processes for handling data, from initial publication to updates and, if necessary, retirement or archiving.

Building on Past Work: Cambridge's Open Data Program has always been conscious of the balance between openness and privacy. Our Open Data Ordinance (enacted in 2015) and subsequent policies set the expectation that data should be open by default unless it falls under privacy, confidentiality, or security exceptions. In recent years, we introduced practical safeguards. We have used anonymization techniques and geomasking (fuzzing or hiding exact location coordinates) in datasets like crime reports to protect individual privacy. The City's legal team and Open Data Review Board (ODRB), comprising City staff and subject matter experts, review new data releases to ensure that privacy and ethical considerations are addressed. These measures have served us well. For instance, Cambridge has released rich public health and public safety datasets without exposing personally identifiable information, by carefully vetting and sometimes aggregating the data.

However, public concern about privacy is higher than ever. Modern computational methods, including machine learning and data linkage, can potentially re-identify individuals or infer sensitive attributes. Additionally, publicly available information like permit addresses can be exploited for scams targeting residents who conduct transactions with the City. These risks mean we must be even more vigilant in our governance approaches. We have heard from community members (especially those with technical expertise) that transparency must be paired with strong privacy safeguards. One expert noted that if open data is handled carelessly, it could erode public trust and harm vulnerable populations. This feedback reinforces our commitment to privacy-preserving openness as a guiding principle.

Goals and Actions: In the next three years, we will bolster our data governance and privacy protections through several initiatives:

- **Enhance Privacy Protection Techniques**: Cambridge will continue to use and improve techniques like anonymization and geomasking and will explore more advanced privacy-enhancing technologies as they become feasible. For instance, we will monitor developments in differential privacy and synthetic data generation, methods that add statistical noise or generate artificial data to protect individual records. These techniques are still emerging, and many of our datasets are relatively small, which makes some privacy techniques challenging. Over 2026–2028, we will conduct experiments with advanced privacy methods on select datasets to evaluate their effectiveness. We will also monitor state and federal regulations around data privacy to ensure full compliance.

- **Strengthen Governance and Transparency**: We will continue to refine our data review processes and make them more transparent to the public. We will consolidate and document these processes more clearly and publish regular updates on our governance practices As AI tools become more prevalent, the Open Data Program Manager will continue serving on the City's interdepartmental AI Working Group, collaborating with the City Manager's Office Strategic Team and other departments on guidelines for responsible use of AI with data. We will also educate the public about potential risks and limitations when using AI with open datasets, including through our regular workshops and communications.

Outcomes: By strengthening data governance in these ways, we ensure that our push for more open data never loses sight of the people behind the data. Success in this priority will mean Cambridge continues to release valuable information while upholding the privacy and rights of individuals. We measure this success partly in the absence of negatives: no major privacy breaches, no erosion of public trust. But we also seek positive outcomes: **Cambridge as a model city for balancing transparency with appropriate privacy protections**, where residents appreciate both the transparency and the care taken to protect them. In short, we will treat data as a public asset that must be managed responsibly—with robust governance, the "nutrients" in our data soil (value, trust, quality) will remain balanced for the long term.

Strategic Priority 3: Engage Consistently with the Public

Overview: This priority is focused on establishing a more predictable and sustainable rhythm of public engagement and communication around open data. In the past, many of our community interactions (hackathons, public meetings, etc.) were one-off events. Going forward, Cambridge will cultivate an ongoing **cadence of engagement**: a regular series of opportunities that keeps the public informed, invites participation, and builds a real community of data users and enthusiasts. By doing so, we aim to nurture a culture in which data is more readily used by residents, and where the City's data efforts are transparent and responsive to public input. In practical terms, this means committing to quarterly newsletters, recurring workshops, and special events, rather than occasional outreach. It also means continually listening to feedback and adjusting our approach to meet community needs.

Building on Past Work: Cambridge has tried a variety of outreach efforts over the years: from hosting hackathons with local universities, to offering data workshops, and collaborating with local teachers and students. We've seen the potential of these efforts. However, our engagement has sometimes been sporadic. Survey respondents noted that they weren't always sure what the Open Data Program was up to, or how to get involved. In the 2025 survey, 75% of participants rated regular public communications and workshops as important for the next phase of our programs. They want to hear from us and learn with us more often, not just during big projects or crises.

A **“little and often” approach to engagement will build stronger relationships** than a few big events. Just as trust in any relationship grows from regular, genuine interaction, we believe trust in the City’s data efforts will grow if we show up predictably and make ourselves available.

Goals and Actions: We will implement a **structured engagement calendar** and foster an open data culture that is welcoming and participatory. Key components include:

- **Quarterly Open Data Newsletter:** Our public newsletter is the cornerstone of our communication strategy and the **backbone of our engagement cadence**. In it, we share news such as newly released datasets, updates on projects, reports and analyses from all City departments, upcoming events, and highlights of how people have used Cambridge’s data. The tone is clear and non-bureaucratic, aiming to demystify data work and celebrate successes. By syncing the newsletter with our other activities (like workshops and events), we aim to keep everyone, from casual data users to City staff, in the loop on a regular basis. Because of the centrality of the newsletter to our engagement efforts, we will pursue subscriber growth through multiple channels—including our website, workshop announcements, and outreach to community groups—and will experiment with different approaches to reach new audiences.

- **Regular Workshops and Training Sessions:** Each quarter, we host at least one public workshop or training session on open data. We alternate formats to be inclusive: for instance, an in-person workshop at the Cambridge Public Library one quarter, and an online Zoom workshop the next. In addition to our basic Cambridge Open Data Workshop, we will pilot sessions focusing on more advanced topics (programmatic use of our API, Data Visualization). One theme we will incorporate into many workshops is the use of AI in working with data, since this has clearly resonated with participants in recent sessions. Not only do attendees learn something new, but it also keeps us on our toes; it has the beneficial side effect of requiring the Open Data team to stay up-to-date on what AI tools can do with our data. Another beneficial side effect of our classes will be documentation: we will record public zoom sessions and post them in the “how to” section of our website, ensuring that instructional materials remain current as the open data portal changes. All workshops will include Q&A and feedback opportunities.

- **“Big Issues” Public Event Series:** In addition to hands-on workshops, we plan to launch a “Big Issues in Data” talk series. Hosted at the Cambridge Public Library or on Zoom, these will be public talks or panel discussions focusing on how data relates to major community challenges and trends. We will invite a mix of voices: academics from Harvard/MIT and local colleges, industry experts or data scientists, City officials, community organizers— anyone who can contribute perspective on the issue. The goal is to connect the dots between our open data and real-world decisions or research, in an engaging, lay-friendly way. These events will also strengthen our partnerships by involving academic and regional partners.

- **Targeted Outreach and Education:** Beyond citywide events, we will maintain and expand targeted engagement with specific groups, especially educational institutions and youth. In recent years, we have introduced AP statistics students at Cambridge Rindge and Latin School (CRLS) to the open data portal, participated in a Data Stories project for Algebra students at the Prospect Hill Academy Charter School, talked to Innovators for Purpose about AI in the City, and arranged a "data field trip" to the Emergency Communications department for Community Charter School math club members. We aim to double down on these efforts and make them more regular—and reach out to other groups as well (perhaps senior centers). The overarching intention is to meet people where they are: bring open data knowledge to different audiences and listen to their specific interests.

- **Feedback Loops and Continuous Improvement:** Engagement is not just about broadcasting information; it's also about listening and improving. We will build in feedback mechanisms for every initiative. Our newsletter will invite readers to respond with questions or suggestions in each. Workshops will have quick feedback surveys at the end to gauge what worked and what didn't. We'll monitor attendance and interest; if certain topics draw big crowds or certain outreach channels prove effective (or not), we'll fine-tune accordingly. Over time, this will help us get to know our public better and focus on the forms of engagement that are most meaningful.

Measures of Success: The expected outcome of this priority is a much stronger presence and relationship between the Open Data Program and the community. Instead of being a behind-the-scenes data provider, the program will be an active, visible participant in civic life. Our goal is that that by 2028, Cambridge will have a network of open data stakeholders, from casual newsletter readers to workshop regulars to civic tech collaborators. This human infrastructure is just as important as the technical infrastructure. When more people are engaged with open data, the City benefits from new insights and contributions, and the public benefits from greater transparency and collective problem-solving.

Not all of this can be measured, but some can: How many participants attended our outreach events? How many feedback responses did we get and how did they rate us? How many subscribers to our newsletter? We will track and monitor and publish and hold ourselves accountable to these metrics.

Strategic Priority 4: Empower City Departments as a Connector and Enabler

Overview Within City government, the Open Data Program will embrace a role as a connector and enabler rather than a direct service provider or "doer" for every data project. Cambridge is fortunate to have talented staff across dozens of departments. Our program's mission is to help those departments use data more effectively by linking them to the right tools, standards, training, data resources, and to each other. In other words, we

will act as a “matchmaker” and a “best supporting actor” for internal data efforts, rather than trying to star in every performance ourselves. This priority recognizes that long-term success in data-informed governance comes from building capacity throughout the organization. By empowering other departments and facilitating data sharing across departments, we enable many more data initiatives to flourish, and they’ll be more sustainable because they have buy-in and knowledge distributed across the City.

This priority may be the most consequential in this plan. While public access to data creates transparency and accountability, the primary way open data improves residents' lives is through better government operations. When City staff across all departments can access, share, and analyze data effectively, services improve, resources are allocated more efficiently, and decisions are better informed. The 2025 survey confirmed this: respondents rated 'helping other City departments share and use data more effectively' as more important than any other priority.

Building on Past Work: In the 2010s, Cambridge (like many cities) initially centralized some data analytics work to jump-start the effort. The Open Data Program partnered with various departments on projects ranging from analyzing traffic data to creating dashboards for public health. Those collaborations demonstrated the value of data, but they also highlighted that departments often need ongoing support and tools, not just one-time help. In recent years, we began facilitating a City Data Analysts Group, an informal network of City employees who work with data in their roles. This group meets monthly to share projects, technologies, and approaches. For example, an analyst from the Community Development Department might present how they use GIS and R for demographic analysis, or an IT staff member might demo a new AI coding tool. These sessions help break down silos, introduce colleagues to each other, and spur reuse of solutions.

We’ve also learned that some of the most successful department-led data initiatives happen when the Open Data Program stepped back from a “doing” role and instead provided guidance or connective tissue. They are the subject matter experts – we cannot possibly approach the depth of their knowledge. What we provide is the platform, the tools, best practice, initial training and ongoing support. That department’s staff can then run with it and maintain the data themselves. They feel ownership, and the data is kept up-to-date because it is integrated into their workflows. The 2025 survey confirmed the importance of internal capacity: respondents rated “helping other City departments share and use data more effectively” as more important than any other priority. To meet these expectations in the most effective way, we will explicitly embrace our connector role.

Goals and Actions: Empowering City departments involves being an internal hub for data knowledge, providing targeted assistance, and fostering a supportive culture for data use. Here’s how we will do it:

- **Maintain and Grow the Internal Data Network:** We will continue convening the **City Data Analysts Group** on a regular schedule (monthly meetings). These meetings (usually over Teams but sometimes in-person) let staff from different departments share what they’re working on, exchange tips, and build relationships.

We will ensure these sessions are welcoming not just to "analysts" in their title, but anyone who touches data as part of their job—planners, finance staff, librarians, etc. Each session might have a spotlight presentation or two, followed by open discussion. The Open Data team will facilitate these meetings, but the content will largely come from the participants. Our role is to connect people to each other and help them learn from each other's successes and challenges.

- **Act as Technology and Data "Matchmaker:"** The Open Data Program will help departments get access to the right tools, training, and data from across the City. We will connect people with technology and guidance, effectively acting as an internal consultant on data needs. A key part of this role is facilitating interdepartmental data sharing—for example, making ISD's permit datasets available to Community Development for tracking housing development and solar installations, or connecting Public Works data with Traffic & Parking for infrastructure planning. When departments need specific tools (such as interactive dashboard platforms), we might help them obtain a license, provide a starter tutorial, or connect them with another department already using that tool. We will inventory the data-related tools and datasets the City has and make sure departments know what's available, avoiding redundant purchases and encouraging reuse. When new needs arise, we can pilot solutions centrally and then distribute them. Training is a big part of this: as new tools or methods emerge, we will arrange short training sessions for staff or point them to existing resources. Our internal trainings and outreach will run a similar (but slightly staggered) cadence to our external efforts.

- **Assist with Standards and Data Management:** As the central open data team, we develop City-wide standards and best practices for data management that departments can adopt. We will encourage departments to incorporate open data publication into their project life cycles (for instance, when a department launches a new system that collects data, part of the plan should be how to publish appropriate data from it). Our team will consult on such projects, making sure data architecture choices don't inadvertently complicate openness or analytics down the road. In essence, we'll act as internal data consultants/standards-bearers, helping departments help themselves. This connector role also extends to linking departments to external data sources: if a department needs data that another City or agency or a university might have, we can broker that connection.

- **Support Data-Informed Decision Making in Departments:** Beyond technical tools, empowering departments is also about culture and process. We will work with department heads and the City Manager's Office to integrate data use into routine decision-making. We aim to create more data champions within departments – staff who are enthusiastic about data and can lead their colleagues by example. Part of our job is to identify and support these emerging champions (often they are already in departments but may not have the recognition or

platforms to amplify their work). Through recognition (like shout-outs in newsletters) and support, we can motivate more staff to step up as data leaders.

- **Avoid Over-Reliance on Central Team:** A candid aspect of this strategy is acknowledging the limits of the central Open Data staff (which is small). We simply cannot execute every analysis or maintain every dataset on behalf of all departments—nor would that be healthy in the long run. By being a connector, we intentionally **shift from a "we'll do it for you" service model to a "we'll help you do it (and do it well)" model**. Of course, we will still directly manage the core open data portal and lead citywide initiatives (and in some cases we might take on projects for high-priority needs), but our primary role will be enabling others. This approach will make our program more resilient; knowledge won't be bottled up in one or two individuals but spread across a network of City staff. It also elevates the data maturity of the whole organization. By 2028, we want Cambridge to have a reputation internally as a place where "almost everyone is a data person" to some degree—not just a specialist team. That may sound ambitious, but incremental progress is achievable: for example, if every department has at least one person who knows how to publish a dataset and one person who can do a basic analysis, that's a great foundation to build on.

Outcomes: This priority's impact will be seen in both culture shifts and concrete indicators: more departments actively publishing their own datasets, growing participation in the City Data Analysts Group, increased cross-department data collaborations, and positive feedback from staff about feeling more supported and capable in using data. Cambridge was one of only two North American cities to be awarded three times with Gold-level certification in Bloomberg Philanthropies' What Works Cities program for our data-informed governance. Building on such recognition would be a powerful validation that we're on the right track. Ultimately, an empowered City workforce means better services and smarter decisions for Cambridge residents, which is the end goal of all these efforts.

Strategic Priority 5: Encourage Responsible Innovation through Structured Experimentation (and Integrated AI)

Overview: The world of data and technology is evolving quickly. New tools, techniques, and ideas emerge constantly, from artificial intelligence breakthroughs to novel ways of visualizing information. Cambridge's Open Data Program must stay agile and forward-looking, while avoiding the trap of chasing every trend. The key is **structured experimentation**: we will encourage innovation by running small-scale pilots and experiments, but in a disciplined way that aligns with our core goals and values. This priority is about creating space for creativity and learning ("the fun stuff," as one staff member put it) without losing focus on our main mission. A major component of this priority is the thoughtful evaluation of what benefits the integration of Artificial Intelligence (AI) might bring to our work. We view AI not as a separate silo, but as a powerful new thread

that can be woven through all aspects of open data—from improving data quality, to enhancing user experiences, to generating insights. Our experiments with AI (and other advanced technologies) will be deliberate and ethical, with a constant eye on accuracy, privacy, fairness, and real-world usefulness.

AI is a Cross-Cutting Theme: It's worth stating up front that AI is not a stand-alone priority in a vacuum; it's part of everything we do. In this strategic plan, we deliberately mention AI across multiple priorities: making data AI-ready in Priority 1, using AI in workshops (Priority 3), considering AI ethics in governance (Priority 2), etc. We do so because if we're approaching AI correctly, its integration will naturally occur in many areas, rather than as an isolated effort. That said, we highlight it here to ensure our approach is clearly articulated, not to add "flash" to our plan. The reality is that **AI is a big deal**—possibly more transformative in this decade than the internet was, or than the pandemic was in catalyzing change. We can't ignore it, so we choose to address it head-on, with thoughtfulness. The following goals outline how Cambridge will experiment and innovate responsibly, with AI as a significant element of that effort.

Goals and Actions:

- **Establish an Experimentation Framework:** We will create a simple internal framework to guide which new ideas or pilot projects we undertake. This framework will act as a filter to ensure experiments serve a purpose and have defined scope. Criteria might include: (a) relevance to our strategic goals (does it strengthen the data "roots," improve engagement, etc.?); (b) clear hypothesis or question (what are we trying to learn?); (c) limited duration and scope (e.g. a 2-3 month pilot, not an open-ended commitment); and (d) evaluation metrics (how will we know if it was successful or worth continuing?). By requiring this structure, we **encourage experimentation but also keep it contained and aligned**.

- **Focus Areas for Innovation:** Based on our strategic focus, we see a few key domains where experimentation will be most valuable:

    - **Improving the "Roots" with Advanced Tools:** Some experiments will be about using cutting-edge technology to enhance our core operations. For example, we plan to develop a prototype **"dataset health dashboard"**—a tool that automatically monitors each open dataset for freshness (last update), completeness, and usage stats. This dashboard might use a bit of AI or statistical analysis to flag datasets that are potentially outdated or under-utilized. Another experiment in this vein is **using AI for automated data quality assurance**: we could pilot an AI system on a subset of datasets to detect anomalies or inconsistencies (as mentioned under Priority 1) and compare its performance to manual QA. The idea is that **tools to make us better at the basics can themselves be quite innovative**. If successful, such tools can be operationalized to continuously support the program (making the experiments very worthwhile).

- **High-Impact Thematic Pilots:** We will identify one or two areas of high importance to the City where data innovation could make a significant difference, such as public health or climate resilience. These experiments will likely involve partnerships with academic researchers, healthcare providers, or civic tech organizations.

- **Open Data as an AI Sandbox:** We consider Cambridge's open data to be a perfect testing ground for new AI tools, precisely because it is open and non-confidential. We can be a **"canary in the coal mine"** for safely exploring how modern AI techniques can be applied in the municipal context. By running experiments on open data (where privacy risk is less sensitive since the data is already public), we can learn lessons that inform other departments' use of AI on more sensitive data later.

- **Advanced Privacy and Ethical Tech:** In line with Priority 2, we will also run experiments with technologies that could enhance privacy or ethics in data. For example, techniques like differential privacy might allow us to publish more granular neighborhood-level statistics (such as health or demographic data) without risking individual identification. Synthetic data generation could enable external researchers to train machine learning models on realistic Cambridge data patterns without exposing actual resident information We might pilot a differential privacy tool on a test dataset to evaluate how much utility is retained versus privacy gained, or experiment with generating synthetic datasets that preserve statistical patterns but contain no real personal data. These experiments would be done in collaboration with experts (perhaps at local universities or the state) and if promising, could pave the way for implementing these techniques more broadly, both in Cambridge and as a model for other cities.

- **Promote a Culture of Innovation (with Guardrails):** Beyond specific projects, we want to cultivate an internal mindset that innovation is welcome—*with responsibility*. City staff should feel they can propose a new idea or try a new approach, especially if it could improve services or efficiency, and know that the Open Data Program will support them. We will create channels for staff to pitch small projects or ideas that leverage data or AI and provide guidance or even modest resources to help incubate them. At the same time, we will ensure everyone understands the "guardrails": for example, any experiment involving personal data must go through privacy review; any use of AI in decision-making must have human oversight, etc. Essentially, we normalize experimentation as part of our workflow, rather than something ad hoc or secret, and we normalize critical evaluation of experiments too. If something doesn't work out, that's fine—we document why and move on, having learned from it. By 2028, we'd like to see a City where pilots are constantly happening at small scale (in traffic, in human services, in finance analytics, etc.), many involving our open data, and the best of those are regularly integrated into full programs.

- **Integrate AI Thoughtfully and Transparently:** Specifically on AI, our commitment is to integrate it **thoughtfully and practically** in service of our objectives. Our approach to ethical AI use is grounded in several core principles: accuracy (AI outputs must be verified and reliable), transparency (we disclose when and how AI is being used), fairness (we actively check for bias and discriminatory outcomes), privacy (we protect sensitive information), and, most of all, **human accountability** (humans remain responsible for decisions, with AI as a tool, not a decision-maker). We will explicitly call out in our project plans where AI is being used, so we can discuss it openly rather than letting it slip in unchecked. We will continue to be involved in the ongoing development of Citywide ethical AI guidelines for the City and data-specific sub-guidelines. Moreover, we will engage the public on AI: through our workshops and Big Issues talks, we'll educate residents on both the potential and pitfalls of using AI with city data.

Outcomes: By embracing structured experimentation, Cambridge aims to remain innovative without chasing fads. We expect to gain valuable insights from our pilots. Some will lead to new features or services, others will teach us what not to do, which is equally valuable. By 2028, we hope that some experimental tools become regular parts of our platform. We also anticipate that our thoughtful approach to AI will position Cambridge as a leader in responsible municipal technology. Success means both concrete innovations and cultural shift: City staff who are curious and confident about trying new approaches, and a public that sees Cambridge as innovative but grounded.

Strategic Priority 6: Strengthen Partnerships in the Broader Data Ecosystem

Overview: Cambridge is not an island when it comes to open data. We are part of a regional and national ecosystem of governments, institutions, and communities working on open data and civic technology. By actively participating in this broader ecosystem, we can accelerate our own progress and contribute to collective knowledge. This priority focuses on forging and deepening partnerships outside of our City department structure: with neighboring cities and towns, with state and federal agencies, with academic and research institutions, civic tech groups, nonprofits, and networks like What Works Cities. We want to share Cambridge's experiences and lessons so others can benefit, and we want to learn from what others are doing so we can adopt best practices faster. By aligning with broader initiatives, our relatively small program can leverage resources and ideas from a much larger pool.

Building on Past Work: Historically, Cambridge has been very active in the open data arena. We were one of the first cities in Massachusetts to have an open data ordinance and portal, and we've collaborated with Boston and the Metropolitan Area Planning Council (MAPC) on data projects. Cambridge was a pilot member of initiatives like Bloomberg's What Works Cities (WWC). Through sustained effort we first achieved WWC Gold Certification in 2021, signaling excellence in using data and evidence in government.

That process itself involved partnership: WWC provided us with expert coaching and a network of peer cities to learn from. We have also engaged in national conversations, with Cambridge staff attending and speaking at conferences on open data and smart cities, bringing back ideas on standards and community engagement techniques. Additionally, our proximity to world-class universities has supported a range of research collaborations and student work that uses Cambridge data to explore emerging technologies and urban issues.

However, we see opportunity to be even more systematic and proactive about partnerships. The 2025 survey asked what partnerships Cambridge should pursue, and respondents suggested closer ties with local universities, civic tech meetups, and regional data sharing on cross-border issues (e.g., regional transit data). One clear message is that Cambridge is expected to be a leader and convener in the open data space, not just focus inwardly. This means choosing partnerships where we can have mutual benefit.

Goals and Actions:

- **Strengthen Local and Regional Collaboration**: We will deepen our collaboration with neighboring communities and regional bodies by setting up regular touchpoints with data teams in cities like Boston, Somerville, and Brookline. We have already begun conversations with neighbor cities about the formation of a Metro Boston Open Data/Analytics Roundtable where Greater Boston municipalities can discuss projects, challenges, and opportunities to work together. Possible outcomes could include shared datasets across city lines (for example, a regional bike lane network dataset), joint events, or coordinated advocacy to state government for support or policy changes.
- **Engage with Academic and Research Partners**: Cambridge is uniquely positioned in a region dense with world-class universities—home to MIT and Harvard and surrounded by institutions such as Northeastern, Boston University, Tufts, and others. We will actively seek partnerships with academia, where faculty and students use the city as a living lab and in return provide insights or tools we can use. This could include research partnerships on city data, capstone projects, academic advisors on working groups, student internships and fellowships, or collaboration on understanding how to apply AI to City data responsibly. These partnerships will ensure we're tapping into the latest thinking while providing a practical testing ground for new theories and tools.
- **Participate in National Networks and Civic Tech Partnerships:** Cambridge will maintain an active voice in national conversations on open data through our relationship with What Works Cities and other initiatives. We will explore partnerships with organizations like the U.S. Conference of Mayors Data Task Force or the Open Data Charter. We will also reach out to the civic tech community and private sector, maintaining relationships with groups like Code for Boston and considering data sharing partnerships with institutions to tackle specific problems.

Where appropriate, we may enter agreements for collaborative projects with these partners.

- **Share Our Work and Advocate**: Many of our experiments will involve partners, allowing us to reduce duplication and share results. We will document and share outcomes of our pilots publicly, contributing knowledge back to the wider community. We will also leverage Cambridge's reputation to advocate for open data best practices and policies at higher levels of government. Cambridge will remain outspoken about the value of open data in democratic governance, standing firm that it is an irreplaceable public contribution that strengthens trust and innovation.

Outcomes: Through these partnerships, improvements that might take us a long time to develop internally can happen faster with collaborators. We foresee tangible outputs like joint datasets, published case studies, or shared tools. One measure of success will be Cambridge's visibility in the open data community: are others referencing our approach as a model? Already, Cambridge's adoption of the PDDL license was noted as clearly articulating our commitment to openness; we want to continue making moves that others cite. Another measure is internal: City leadership and the public should see that our collaborations bring concrete value. For example, if partnering with Boston yields a better way to share housing data, that helps our policy decisions. If partnering with MIT yields a new tool to optimize energy usage in city buildings, that saves taxpayer money. Ultimately, these partnerships should help Cambridge stay current with evolving best practices and contribute our innovations back to the community. By 2028, we hope Cambridge will be both learning from and contributing meaningfully to the networks that define the future of open data and civic technology.

Metrics for Tracking Progress (2026–2028)

This section summarizes the quantitative indicators the Open Data Program will use to track progress on each strategic priority. These metrics help us see whether our efforts are moving in the right direction: Are the foundations of our data ecosystem strengthening? Is our engagement reaching more people? Are our governance and innovation practices functioning as intended? We will monitor these indicators regularly and report on them publicly. While we will establish specific annual targets for some metrics, the primary purpose is to understand trends, identify where we need to adjust our approach, and maintain transparency about our progress. Metrics are grouped by priority for clarity.

Priority 1: Strengthen the Data Ecosystem

- Total number of public datasets
- Dataset downloads/views
- Datasets with automated refresh pipelines
- Dataset freshness percentage (percent updated on schedule)
- Datasets with PDDL license

Priority 2: Strengthen Governance and Protect Privacy

- Datasets with documented privacy rules/geomasking methods
- Privacy inquiries and requests received
- Published governance updates, standards, or guidance pieces
- Public attendance at Open Data Review Board meetings

Priority 3: Engage Consistently with the Public

- Public newsletter issues sent and subscriber count
- Public workshops and attendance
- Targeted outreach events and attendance (schools, community groups, and other partners)
- "Big Issues" talks and attendance

Priority 4: Empower City Departments as a Connector and Enabler

- City Data Analysts Group (CDAG) meetings and attendance
- Departments represented in CDAG
- Active internal open data portal accounts
- Staff trainings and attendance

Priority 5: Encourage Responsible Innovation through Structured Experimentation (and Integrated AI)

- Evaluations of emerging tools or methods
- Documented pilots or experiments conducted
- New tools or methods adopted/deployed

Priority 6: Strengthen Partnerships in the Broader Data Ecosystem

- Academic data collaborations (research consultations, student or faculty projects)
- Civic tech collaborations (Code for Boston sessions, civic hackathons using Cambridge data)
- Participation in regional or national convenings (workshops, summits, community calls)
- Shared outputs with regional partners (shared tools, co-developed guidance)
- External references to Cambridge's work (mentions in What Works Cities, GovEx, USCM, Open Data Charter, etc.)

Conclusion and Call to Action

This Strategic Plan for 2026–2028 charts a clear course for Cambridge's Open Data Program: focus on the fundamentals (our data "soil"), strengthen trust through good governance, engage our community regularly, empower our colleagues internally, innovate

carefully, and collaborate widely. It is a plan rooted in realism and responsibility, mindful of past lessons and current challenges, yet optimistic about what is possible when a city and its people share data openly and use it thoughtfully.

We want to conclude with an invitation and a reaffirmation.

First, an invitation: We invite all our stakeholders (residents, City employees, students, businesses, regional partners) to join us in bringing this plan to life. We welcome your ongoing feedback and ideas. Tell us what datasets you need, what isn't working, what new features you'd like to see, or how you've used Cambridge's data in your own work. You can reach out via our Open Data Portal, email (opendata@cambridgema.gov), or come talk to us at the workshops and events we'll be hosting. Your input will continue to shape our direction, just as the 2025 survey helped shape this plan.

Second, a reaffirmation: We affirm our commitment to the core principle that data in service of the public is a public good. Our mission remains what it has always been: to help Cambridge leverage data to improve transparency, efficiency, and innovation for the benefit of everyone who lives, works, or visits here. While the tools and trends evolve (today it's AI, yesterday it was "big data"), our fundamental dedication is unchanged. We will use data to serve the public good, and we will do so openly and ethically.

In practical terms, this plan will guide our day-to-day work. When deciding how to allocate limited time or budget, we will use these strategic priorities as our compass. They will help us say "yes" to the work that matters most and, importantly, to say "no" to things outside our strategic focus. We will periodically report on our progress so that the public and City leadership can hold us accountable. If we fall short in an area, we'll be candid about it and adjust course.

Cambridge's open data journey continues. By nourishing the soil of our data ecosystem, we expect to see robust growth: more community-driven insights, more data-informed decisions, and perhaps innovations we can't even predict yet. Stay engaged with us. Subscribe to our newsletter, come to an event, publish a visualization or analysis of Cambridge data, or simply drop us a line with a question. With your help and participation, this plan will deliver practical benefits to our community every day.

Thank you for reading, and we look forward to working together to make Cambridge's data ecosystem richer, more accessible, and more productive than ever.