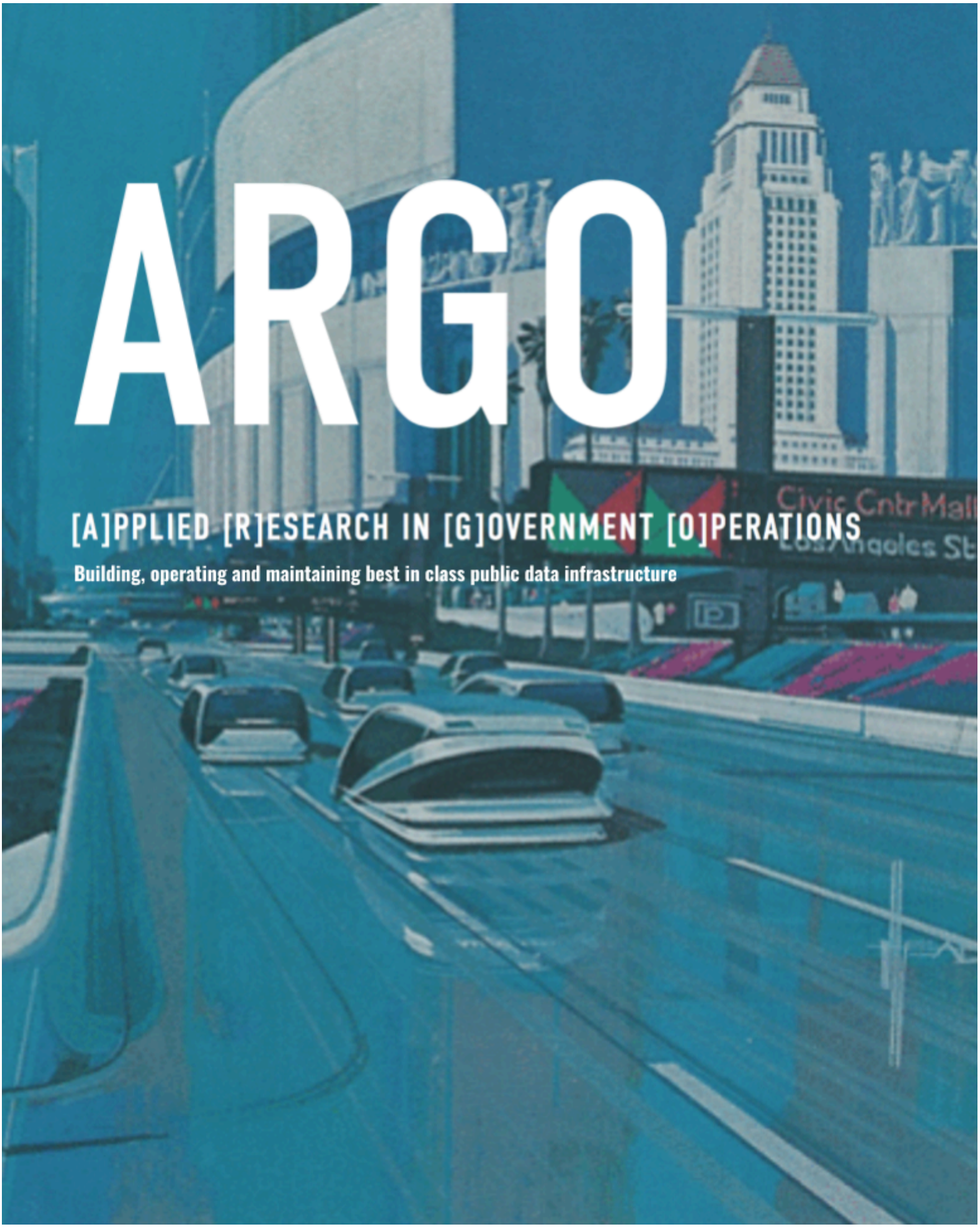


Image Source: LA Times Magazine's vision of the region in 2013 published April 3, 1988 (http://documents.latimes.com/la-2013/)

Introductory Letter: California's Golden Opportunity

To lead the world in honest and effective government with best in class public technology.

From the world's largest port to the nation's first freeways to pioneering aqueducts, California's history has been defined by visionary public works. The digital revolution has transformed countless industries, yet by and large a time traveler from the 50's would find the operational practices of California government strangely familiar.

Nothing equal to the development over a century ago of professional water utilities or universal public schooling — institutions that implemented nearly ubiquitous access to clean water and essentially eradicated illiteracy in America — has been developed for the digital era. California can and should lead the world in building the great public works of our era – public technology that tackles our big challenges as a state.

This stems from a belief that California's open, inclusive values are more important than ever in today's tribalistic, nationalist times and that California is ideally poised to pioneer digitally native government operations. California today faces massive challenges with infamous traffic, increasingly unaffordable housing and the generational task of addressing climate change.

Public data infrastructure will not "solve" those problems. Yet with over 482 cities in addition to the several thousand special districts and local governments in the 58 counties that make up California, it is hard to imagine a future whereby California meaningfully tackles those challenges without underlying data integration.

To achieve that goal, ARGO brings a proven team and battle-tested public data platform which we call *Kraken* to reflect the challenge of taming "wicked" government IT problems. ARGO's team has played key roles in multimillion dollar smart city and data infrastructure projects as part of the federally funded National Resource Network,[1] New York's New Lab,[2] New York University's Data Facility,[3] and the Bloomberg Philanthropies/Aspen Institute Initiative on Autonomous Vehicles.[4]

---

[1] National Resource Network. Hidden in Plain Sight: Why California's Economically Challenged Cities Matter. Varun Adibhatla served as data and research lead. (http://www.nationalresourcenetwork.org/en/Document/306226/Hidden_in_Plain_Sight_Why_Californias_Economically_Challenged_Cities_Matter)

[2] Urban Technology Growth Hub, New York Economic Development Corporation, RFP. Varun Adibhatla was the founding Director of the Urban Tech Hub at New Lab. (http://bit.ly/nycedc_urbantech_rfp)

[3] NYU CUSP Data User Facility, Alfred P. Sloan Foundation. Graham Henke served as platform architect and data engineer.(https://sloan.org/grant-detail/7128)

[4] Bloomberg Aspen Initiative on Cities and Autonomous Vehicles. Varun Adibhatla served as data and research lead for Q1&2 2017. (Autonomous Vehicles: Future Scenarios) (http://avfutures.nlc.org/)

In California, ARGO's water data work has won the best urban technology in the California water data challenge put on by the Obama White House in 2016 in addition to other accolades and accomplishments detailed in the proposal. ARGO is deeply committed to the mission of modernizing government operations with best in class public data infrastructure.

On a personal note, as a fourth generation California native, I am excited by California's public technology ambitions because I believe in its importance for the future of the place I call home. Similarly, the team of civic data scientists that make up ARGO come from around the globe and are committed to ensuring that this platform provides a public service for California that can be replicated in their own home towns.

Cities around the world are investing in public technology initiatives to meet similar challenges, and the public data approaches and technologies pioneered in California can be shared globally. Likewise, California can learn from and redeploy innovations pioneered elsewhere.

As America's gateway into Asia and crossroads with Latin America, I believe California stands poised to lead the world boldly and confidently into the future. The digital revolution offers the ability to connect humans across the globe in ways scarcely imaginable a generation ago, and California's visionary commitment to better public technology offers great potential to utilize those new tools to better tackle some of humanity's oldest challenges.

In delivering that work, ARGO offers battle-tested public data infrastructure, a uniquely nonprofit shared services business model and a foundational commitment to public service.

Cheers,

Patrick Atwater, Q1 2019
patrick@argolabs.org | 213.358.6500
Applied Research in Government Operations ("ARGO")

"You need to have a public sector commensurate with the vitality and creativity of private life here. Look at the universities, the hospitals, our doctors! Look at the vitality of our literary culture, our painters!"

-Kevin Starr, Former State Librarian of California

TABLE OF CONTENTS

ARGO

Accelerating Reform in
Government Operations

The Need for Public Data Infrastructure

To support more efficient, effective and imaginative public service delivery.

Similar to water or electricity, ARGO believes public data is a valuable public resource that deserves technical excellence and a foundational commitment to public service.[5]

That idea of a "data utility" may seem strange, yet remember that little over a century ago, publicly owned and professionally managed water utilities were a radical innovation. Many other municipal artifacts like regular budgeting to rigorously measure the dollar inputs into public services were massive shifts in public management. Today we face a similar frontier as the public sector works to adapt to the digital revolution. One emergent consensus is that this frontier offers the potential for the routine, rigorous measurement of the output public services deliver.

Code for America Founder Jennifer Pahlka wrote and spoke about that "Delivery Driven" approach to government operations and policy at the recent 2018 Code for America Summit.[6] The leader of NYU' Data Facility Julia Lane also writes and speaks regularly about the new frontier in measuring the effectiveness of public service delivery.[7] While still nascent and somewhat in flux, that vision for routinized evaluation of government service delivery represents a transformational shift in how public services are managed.

NYU's GovLab estimates that only $1 out of every $100 spent by governments in the US is backed by rigorous evidence.[8] Over the past seven years, ARGO has worked with leading local water utilities and other partner local governments to scale a conceptual data-value pyramid to deliver actionable decisions. Those water data wins are detailed the next Section "ARGO's water data achievements."

The need for better water data illustrates the larger issues facing the public sector. The last twenty years in California have been the driest in the last eight centuries. That unprecedented

---

[5] See the following Civicist post on why we should manage public data like water: https://civichall.org/civicist/public-technology-movement-can-learn-californias-water-industry/.
[6] Code for America Blog "Delivery-driven Government: Principles and Practices for Government in the Digital Age".(https://medium.com/code-for-america/delivery-driven-government-67e698c57c7b)
[7] Julia Lane "A new frontier: How big data can improve government services". (https://cityandstateny.com/articles/opinion/a-new-frontier-how-big-data-can-improve-government-services.html)
[8] "The GovLab Index: Measuring Impact with Evidence". (http://thegovlab.org/govlab-index-measuring-impact-with-evidence/)

reality will only grow worse with climate change. Meanwhile our water management system is a tangled mess of over a thousand local water retailers and additional layers of regional, state, and federal agencies. That institutional fragmentation means it is shockingly difficult to answer basic questions about how much water is used where, when and to what purpose.[9]

Regions across the globe similarly face climate challenges. That seismic shift is compounded by the additional adaptation challenges of evolving legacy institutions to meet the new reality. That challenge is generally compounded by additional layers like energy geopolitics, aging infrastructure and demographics, democratic instability -- what Adam Tooze eloquently encapsulates in the term polycrisis.[10] There are numerous examples of cities leading the way in implementing adaptive, digitally native operational practices.[11]

ARGO believes state and local governments in California (and beyond!) can be world leaders in this data centered approach. Scaling the data pyramid not only represents the journey of those organizational units but also represents the journey of individual employees and its stakeholders in better identifying where they are on the value pyramid and where they need to be to realize success. The first step is to identify concrete use cases with stakeholders, to determine what data is necessary to collect or integrate from existing sources. Those use cases delineating specific decisions are important to prioritize, clarifying what data to focus on and where it is worth investing resources scaling the data value pyramid.

Scaling the data-value pyramid

[9] Charles Fishman, New York Times: "Water is broken. Data can fix it."
[10] Adam Tooze, Chartbook: "Polycrisis."
[11] Harvard's Data-Smart City Solutions has a nice catalog of civic data use cases, including examples.

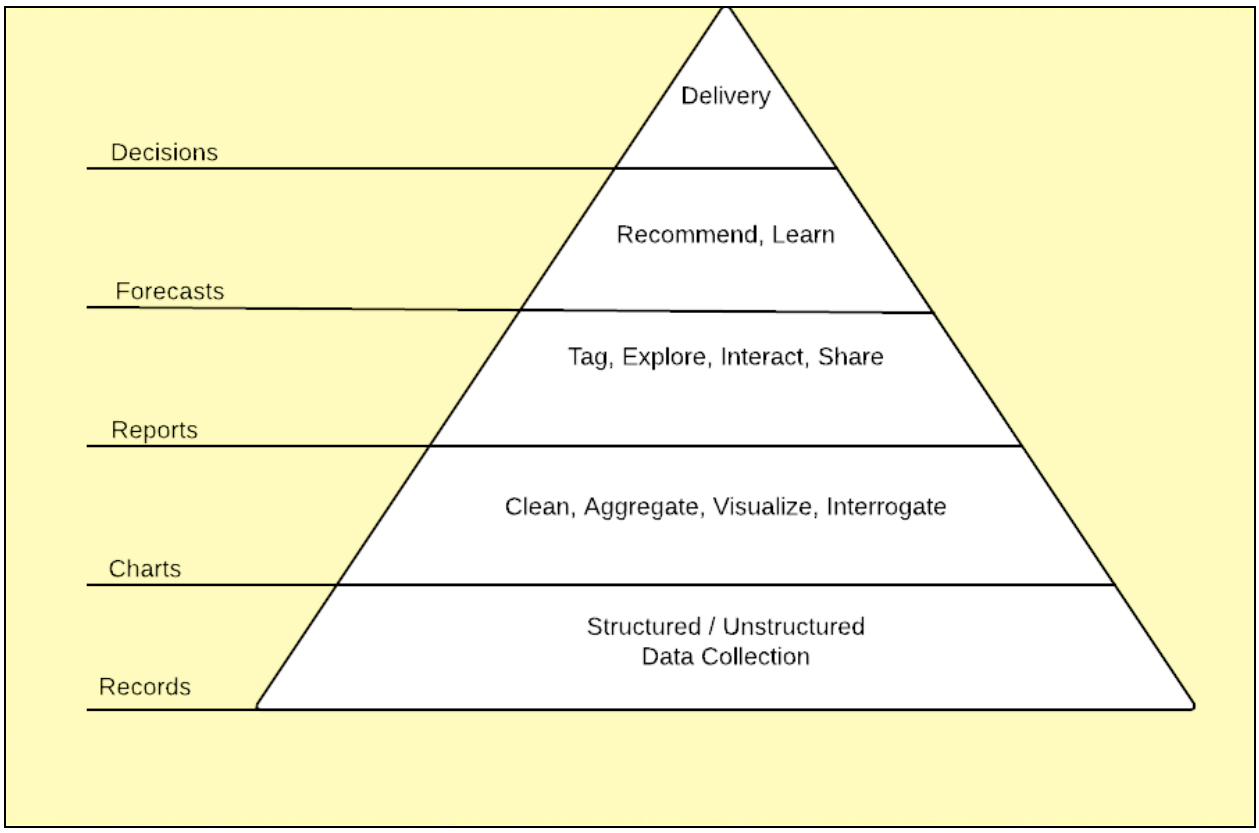


Figure 1: The data-value pyramid.
Source: Agile Data Science 2.0, O'Reilly publications.

Foundationally, the data value pyramid starts at the bottom with data collection that encompasses structured data in the form of data stored in a specific, predetermined though not always standardized format. Examples range from Census data, to data from the Federal Department of Transportation, to local snapshots of community health data. Unstructured data represents data that is collected in real-time through sensors and other instrumentation with the intent to integrate with structured data to deploy rapid and targeted studies and experiments.

Further up the value chain, a discovery component allows stakeholders to easily find the right data they are looking for and to be able to integrate this data with other datasets possibly from disparate domains. As data is collected, aggregated, and discovered, best in class public data infrastructure should be able to power higher-level abstractions such as forecasting and planning tools to recommend specific courses of action for California public sector managers.

Finally, the routinized delivery of distinct and fully formed data products that solves specific problems forms the epitome of the value pyramid. The goal of this data work again is to tackle specific land use, transportation, sustainability and other important public problems in the real

world -- not to simply manipulate data for the sake of manipulating data. That approach has been proven over the past few years in collaboration with leading California water utilities.

ARGO's Water Data Achievements

Governor Brown's recently enacted legislation (AB1668 and SB606) to make "water conservation a way of life" creates an unprecedented need for integrating water and land use data.[12] Specifically, that project utilized parcel level residential land use in combination with landscape area, population and evapotranspiration data to set a residential water budget for each local water retailer across California. That water budget creates an efficiency goal that is used to benchmark residential water demand available in the monthly supplier report submitted to the State Water Resources.[13] A screenshot of the tool is shown below:

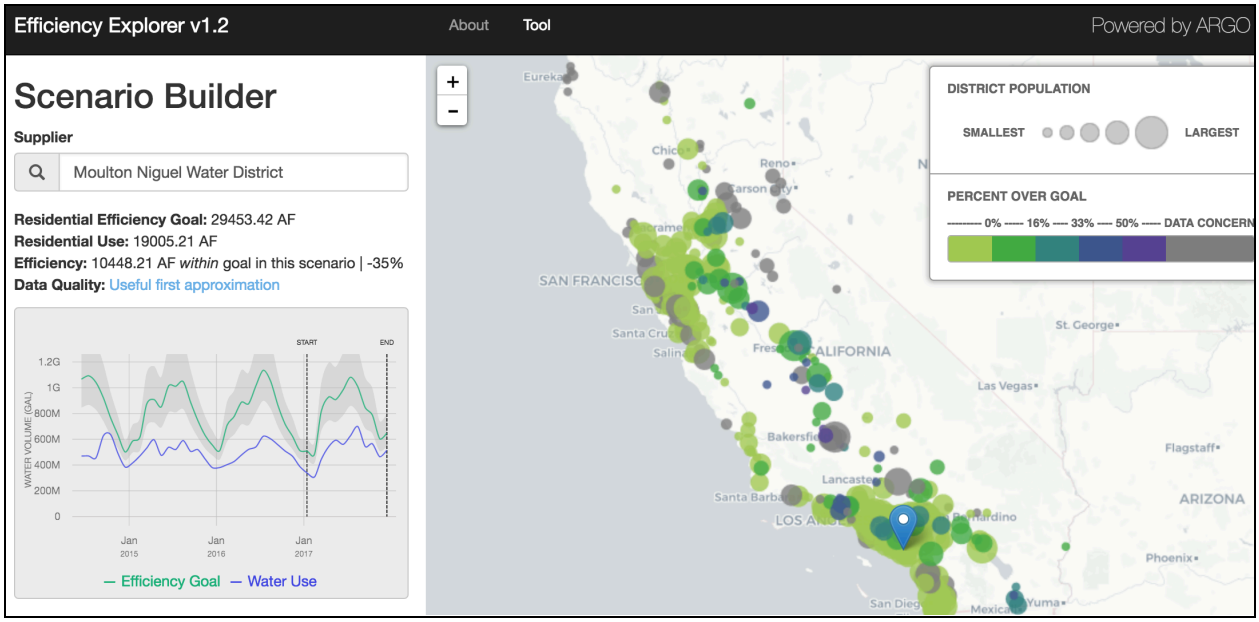


Figure 2: Efficiency Explorer Scenario Builder. See the live tool at FutureOfWaterManagement.org

[12] The importance of land use data is detailed in the following blog post from ARGO's California Data Collaborative water data project website: (http://californiadatacollaborative.org/blog/2017/6/8/residential-water-efficiency-and-the-california-data-quality-landscape).

[13] The full methodology of this initial rapid assessment of the Governor's framework for local water retailers across California is available at: (http://californiadatacollaborative.org/blog/2017/4/28/cadc-statewide-efficiency-explorer-methodology).

That initial data integration, all powered by ARGO's *Kraken* platform, provided the world's first estimates of the Governor's framework for all of California's urban areas.[14] Implementing that framework forms the foundation for ARGO's integrated vision for transforming water management that has been featured in the Journal of the American Water Works Association,[15] Bloomberg Data for Good[16] and regularly updated as part of our annual water data summit.[17] In addition, that water data work has been honored by the Obama White House,[18] Senator Hertzberg's Subcommittee on Modernizing Government, and featured in numerous media publications from the Associated Press to KNX News Radio to Harper's Magazine. In addition, that water data project has also saved partner utilities over $20 million.[19]

The data work on the Governor's water conservation framework also illustrates the power of ARGO's *Kraken* public data platform as described in Sections 3. That data work does not exist in isolation. ARGO is deeply integrated into the fabric of California civic life and is currently powering the data work for the Southern California Water-Energy Task Force on implementing Governor Brown's water conservation framework. California water management has historically been siloed apart from transportation and land use policy decisions. Yet like many other urban challenges such as sustainable land use development and the historic shifts in urban mobility, water increasingly necessitates inter-sectoral institutional collaboration to achieve meaningful solutions. California has the opportunity to build on the water community's historic investment in a new public data platform.

ARGO believes that proven experience delivering this water data work and the depth of civic connections in California uniquely qualifies ARGO to build, operate and maintain public data infrastructure. Further ARGO offers a first-of-its-kind nonprofit shared services model to provide data infrastructure to local government partners. This unique institutional structure enables ARGO to achieve much better results than traditional private sector technology vendors by enabling deep integration and alignment of incentives between ARGO and our partner governments. That deep partnership enables ARGO to cut through the institutional barriers that have traditionally frustrated innovative government data projects and allows a

---

[14] That planning tool is further described in Section 2.5 and can be accessed at (https://futureofwatermanagement.org/).
[15] Journal of the American Waterworks Association: "Adaptation, Integration, and Connection: How Water Utilities Can Plan for Uncertainty". (https://awwa.onlinelibrary.wiley.com/doi/abs/10.5942/jawwa.2015.107.0087)
[16] Bloomberg Data for Good Exchange "Transforming How Water is Managed in the West". (https://arxiv.org/pdf/1609.08715.pdf)
[17] The third annual water data summit will be at USC August 23rd and 24th. See (CAwaterDataSummit.org) for more details and to register.
[18] The Obama White House, Commitments to Action on Building a Sustainable Water Future. (https://obamawhitehouse.archives.gov/sites/whitehouse.gov/files/documents/White_House_Water_Summit_commitments_report_032216_v3_0.pdf)
[19] Moulton Niguel Water District Wins Golden Hub of Innovation Award. (https://www.mnwd.com/mnwd-wins-golden-hub-of-innovation-award/)

pioneering approach since upside reward is aligned with the public interest. ARGO has articulated its approach to public data collaborations in Section 2 of the Technical Approach.

In achieving those larger aspirations, ARGO's team brings a unique set of interdisciplinary experience leading smart cities initiatives across the globe. ARGO has been contracted to deploy its public data platform to support everything from analyzing the impact of congestion pricing in New York City to deploying state of the art sensors to measure street quality conditions to analyzing California's most economically challenging cities. ARGO has already been organically developing foundational open source tooling in line with California's goals (including, but not limited to, the gas tax funding for local street maintenance), which leverages Open Street Map as part of an integrated approach to plan and analyze autonomous vehicle scenarios. ARGO believes those pro-bono collaborations could not be more aligned with California's goals and illustrate that ARGO is fully committed to the mission of modernizing government operations.

One would be hard pressed to find an organization that can match ARGO's institutional commitment to public service, proven public data platform, pragmatic ethos of ongoing public sector experimentation, global breadth of public data expertise and depth of experience in California.

ARGO's Technical Approach

To building, operating and maintaining public data infrastructure

This portion of ARGO's plan is subdivided into three sections:

1) A narrative discussion of key risks and opportunities with public data infrastructure.
2) An overview of ARGO's approach to public data collaboratives to address the issues in section one.
3) A detailed description of ARGO's Kraken platform that powers those public data collaborations.

Section 1: Risks and opportunities with public data infrastructure

California's ambitious public technology goals present a significant opportunity for the state. However, like any complex system integration project involving many stakeholders, there are substantial risks and challenges in developing public data infrastructure. California needs to strategically address several foundational questions:

- First, what is the strategy for dealing with vendor lock-in?

- Second, how will this platform be sustained over the long-term?

- Third, how will this platform avoid the long history of "smart cities" projects that had high expectations and suboptimal implementation?

ARGO believes the answer to these three questions are interrelated and are primarily institutional rather than technological.

Technology vendor lock-in presents a perennial challenge for public sector and large enterprise organizations. In particular, any project of this size, scope, and complexity presents significant procurement challenges and the unique nature of public data infrastructure requires innovative approaches to address those challenges. At a foundational level, public data infrastructure should be developed utilizing open source and well-documented code wherever possible in accordance with established security protocols to enable other firms, academics, local governments and the civic technology community to build upon or refactor code. ARGO utilizes open source by default, and has found this practice helps avoid private vendor lock-in by enabling new entrants to more easily build upon past work.

Furthermore, while properly functioning public data infrastructure should enable many vendors to deploy technologies, the underlying platform needs to be built, operated and maintained in an integrated manner with appropriate oversight. The alternative is system integration consultants managing other system integration consultants and skyrocketing costs -- an all too common tale in government technology projects. ARGO has proven experience navigating those issues in our water work.

In addition, any vendor that delivers phase one of the project will be invaluable in future work. The mission-critical question then becomes: *How will California ensure that public data infrastructure does not become a steady revenue stream for rent-seeking private interests?* ARGO proposes addressing the lock-in issue head-on by aligning institutional incentives. ARGO operates as a quasi-governmental nonprofit that builds, operates and maintains public data infrastructure to support local governments in tackling common challenges.

Through its first-of-its-kind California Data Collaborative ("CaDC") project, ARGO has worked with water utility stakeholders across California to develop a steering committee among participating agencies to regulate annual subscription fees and ensure aligned incentives among all involved. The CaDC steering committee policies are available on the CaDC website here (http://californiadatacollaborative.org/steering-committee/).

This deep institutional alignment also clarifies the long-term sustainability and growth of public data infrastructure in California and beyond. ARGO proposes utilizing the same "data collaborative" co-op business model with subscription fees regulated by participating local municipalities to support long-term maintenance and operations of public data infrastructure.

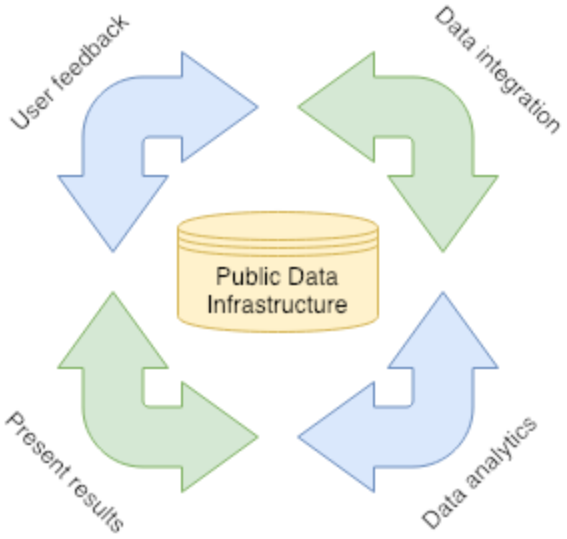


Figure 3: ARGO's iterative, collaborative approach to standardizing data, developing analytics, presenting results and incorporating user feedback.

These data collaboratives would be developed around local governments pooling data in order to tackle common challenges. This model enables economies of scale to offer a much lower cost of service for participating agencies as well as access to a leading global network of civic data science and academic research talent. Potential data collaboratives are further detailed in Section 2 "ARGO's Approach to Public Data Collaboratives."

ARGO's data collaboratives are focused around aspirational challenges that the group of local governments will address. For example, the California Data Collaborative coalition of water utilities is working together to ensure water reliability no matter what the future holds. ARGO believes this simple act of clarifying a common vision helps avoid having a tool in search of a problem, a common cause of smart cities failures.

Furthermore, the data collaborative model is premised on the idea that local government staff have invaluable domain expertise that is instrumental to the success of the project. That institutional knowledge is key at all stages of the technology development lifecycle, from aligning on a standard schema, to developing data transfer protocols, to testing use cases for new dashboards or reporting tools and iterating through the entire analytical life cycle.

Furthermore, that entire data pipeline is only the start of the much longer process of translating the work into improved public outcomes. The headline on the CaDC website states "The Future of Water Management": ultimately it's not about the data so much as the improved public management that new data tools enable. That fundamental belief in focusing on the underlying public challenge and letting technology, analytical and other technical decisions

flow from what will best address that challenge, pervades ARGO's technical approach to building, operating and maintaining public data infrastructure.

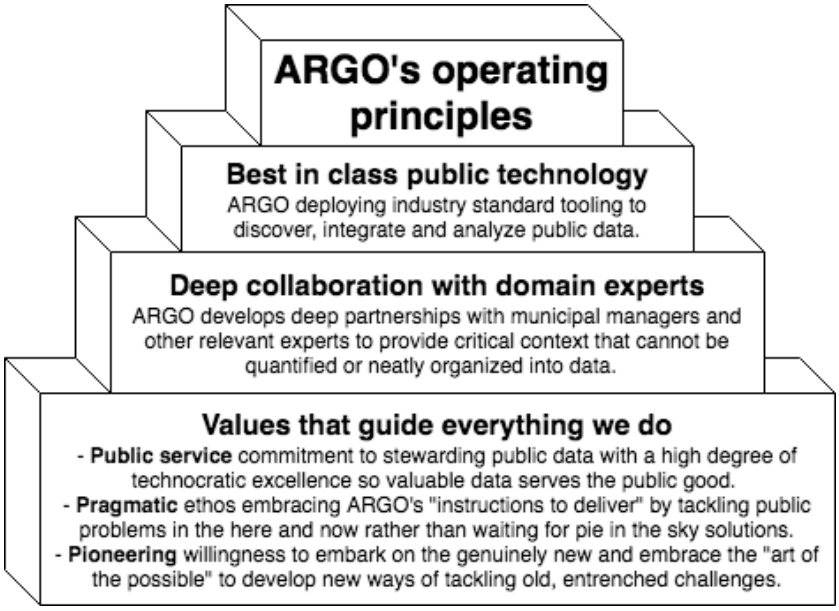


Figure 4: ARGO's principles that guide how we deliver public data work and achieve our mission of supporting local governments in navigating a time of incredible change.

Section 2: ARGO's Approach to Public Data Collaboratives

ARGO believes California has the golden opportunity for a public technology breakthrough. Public data integration provides the foundation for more robust planning efforts and can also enable additional opportunities to support cities in optimizing their operations. For example, New York City was able to improve their return on inspection man-hours by 5x by centralizing and leveraging operational intelligence from departments across the city.[20] ARGO has already helped its local utility partners save over $20 million in avoided capital costs.[21]

Many cities lack the resources to implement an open data program, let alone any sort of advanced analytics. Almost a decade after President Obama's Executive Order declared federal data to be open by default, less than one percent of cities across the nation have an open data portal.[22] This reality calls for shared services across local governments, a model ARGO has successfully pioneered in California's water industry and which others working deep

[20] O'Reilly media: "Predictive data analytics is saving lives and taxpayer dollars in New York City". (https://www.oreilly.com/ideas/predictive-data-analytics-big-data-nyc)
[21] "MNWD Wins Golden Hub of Innovation Award". (https://www.mnwd.com/mnwd-wins-golden-hub-of-innovation-award/)
[22] Calculated based on numbers available from the Open Knowledge Foundation Census of US City Open Data Portals and the National League of Cities.

in the trenches of government operations reform have seen a need for. (See the Open Government Coalition, Shared Streets and Mobility Data Specification).

Globally there has been a Cambrian explosion in "data collaboratives" that bring together organizations across sectors to pool data and deploy open-source analytics to improve public service delivery.[23] Organizations, primarily private companies, which offer ride sharing, autonomous vehicle networks, business entity resolution and other new technologies offer incredibly valuable data for urban planning and local government operations. The data collaborative model of pooling data across the public and private sectors thus offers a great number of strengths within the rapidly changing data technology landscape.

ARGO utilizes this model in developing and managing data collaboratives to advance the public interest. The following subsections provide a high-level overview of ARGO's approach to developing data collaboratives and a detailed description of a nascent ARGO data collaborative focused on city streets (StreetsDataCollaborative.org).

2.1 Public Data Collaborative Development

ARGO takes an analytically rigorous approach to developing new data collaboratives and has scoped opportunities across several local government domains. Based on ARGO's experience and extensive research into global best practices, ARGO looks for local leadership to champion a project, alignment with state policy priorities, and staff capacity to deliver a project before initiating an engagement. If all those criteria are met, ARGO works to dig deep into the specific problem by investigating existing data, qualitatively reviewing relevant academic research and conducting interviews with service providers, city management, service users and other key actors.

That initial inquiry and humility in understanding the institutional context is invaluable in launching a new data collaborative. Much of the most valuable government data exists in either illegible formats or is not shared publicly. Trust needs to be built with partners at the state and local level. That can be done by sharing best practices in other domains and also

There are a wide variety and opportunities to create value through a data collaborative. ARGO's water data work pooled data across participating water retailers. There's also opportunities for public-private partnerships and creatively utilizing private tech sector data. The Mobility Data Specification also highlights the opportunity for a collaborative approach to a data standard and how that can be specified into an API. Gov Lab articulates and provides examples for each

[23] See DataCollaboratives.org for an overview of global initiatives from NYU GovLab.

of those models.[24] The specific type of data collaborative to be deployed depends on the particular needs of the situation.

For example, ARGO has worked deeply with water financial information and identified opportunities to improve standard local government Consolidated Annual Financial Reports ("CAFRs").[25] [26] ARGO has also prototyped tooling to support a more clear operational picture of available homelessness services in Boulder and Seattle.[27] Those were developed in collaboration with leadership at direct service organizations in the areas and offer a potential opportunity given California's tragic homelessness crisis.[28] Lastly, ARGO has worked with Chief Data Officers and Public Works Departments across the country on data issues relating to city streets and has articulated a potential Streets Data Collaborative described in greater detail in Section 3.2.

These examples hint at the tremendous opportunity to transform how local governments operate. Digitally native public services are increasingly in demand as a tech-savvy younger generation expects public services to provide the same quality of user experience they receive in other areas of their life.[29] Within that epochal shift in public administration, ARGO believes firmly that public problems are most effectively addressed through a combination of top-down

---

[24] Ibid.
[25] The New Public Administration Paradigm. ARGO (https://medium.com/a-r-g-o/the-new-public-administration-paradigm-2c1733770365)
[26] Twitter exchange with Joel Natividad, Director of Open Data, Open Gov Inc.(https://twitter.com/vr00n/status/983765912791678976)
[27] ARGO's homelessness data and public technology initiatives include:
- Scaling out (Upswyng.org), a low-cost, open-source, mobile ready alternative to 211.
- YIMBY Stat - Integrating local parcel data with census data to identify areas in a city that are more amenable to YIMBY programs - See our deployment for Seattle based: (github.com/argo-marketplace/Block-Project).
- Housing Voucher mapping and explorer.
    - Mapping housing projects that accept Housing vouchers to help recently awarded vouchers find a home quickly. NYC example: (nyu.carto.com/u/varun-cusp2/builder/cb8f9d3b-995b-4db2-a7d5-848e4fe39635/embed)
    - Developing decision support systems for service providers to quickly identify vouchers based on eligibility. (bit.ly/VoucherExplorer)
- Operationalizing insights from the Campbell review: "Effectiveness of interventions to reduce homelessness" - that aggregate 43 randomized controlled trials.This review represents the least assailable research on homelessness reduction. (https://campbellcollaboration.org/library/effectiveness-of-interventions-to-reduce-homelessness.html)

[28] ARGO has also been deep in the weeds on California parcel level land use data quality issues as part of our work on Governor Brown's framework to "make water conservation a way of life," a key data point in informing permitting decisions and potential policy changes related to new housing to address California's underlying housing affordability issues.
[29] Techcrunch "Millennials could be the answer to government malaise."(https://techcrunch.com/2018/04/04/millennials-could-be-the-answer-to-government-malaise/)

and bottom-up forces. The data collaborative model combines both the tremendous power of technology at scale and the critical local context of on the ground domain expertise.

2.2 Potential Case: Pioneering a Streets Data Collaborative

Regional infrastructure consists of a meshwork of state and local interests. Across the nation, local governments are responsible for maintaining and improving 3.18 million miles of road or 77% of the total 4.12 million miles of road in the United States.[30] Our work in street maintenance has shown that local and regional infrastructure has lacked a modern analytical approach as employed in other, more visible, domains of local government. Local and regional debates about infrastructure development remains a top issue.

[30] American Road and Transportation Builders Association. (https://www.artba.org/about/faq/)

Here lies a key opportunity for California to:

- Support cities across the state to maintain their existing infrastructure and better forecast infrastructure capital expenditure by leveraging low-cost sensors, and cloud infrastructure for frequent digital street surveys.
- Rapidly deploy low-cost, real-time bus tracking for its member agencies to improve public transportation usage and experience.
- Serve as a neutral data intermediary between public agencies and autonomous vehicle manufacturers / ride sharing companies (Waymo, Uber, Lyft, or Tesla) to collaboratively develop strategies to provide mobility solutions targeting populations who most need them.

These three initial opportunities are further described below.

Adopting a digital Pavement Condition Index using SQUID

California cities and counties face a $78 billion shortfall[31] over the next decade to adequately maintain the existing network of local streets and roads, yet approaches to understanding the conditions of our street networks remain a relic of the past.

Today, data collection for resource allocation and performance evaluation exists at two unsustainable extremes. At one end of the spectrum, cities rely on the low-touch, windshield surveys based on the Pavement Condition Index (PCI).[32] This expensive, post-hoc, 1980s-era manual data collection approach provides little ground-truth, and attempts at network-wide collection are inevitably constrained by human subjectivity and fatigue.

At the other end, cities have turned to high-touch, laser and lidar vans. But this ground-truth comes with a price tag. Just recently, the LA's Bureau of Street Services has spent, over a three-year term, $1,428,427 on a private consulting firm to collect data on pavement distresses across LA's 5,500 center-lane miles of streets.[33] This resource intensive and costly operation results in low-frequency surveys of LA's street infrastructure.

Instead of the above two options, California could embrace advances made in machine learning, low-cost sensors, and the wide spread of open-source technology to empower cities like LA and others to collect complete street quality surveys in weeks and perform these surveys frequently enough to anticipate which streets need maintenance first. Instead of

[31] California Senate Bill No. 1 (SB1). (https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180SB1)
[32] Shahin, M. Y., M. I. Darter, and S. D. Kohn, Development of a Pavement Maintenance Management System, Volume I: Airfield Pavement Condition Rating, AFCEC-TR-27, Air Force Civil Engineering Center (AFCEC), November 1976. (http://www.dtic.mil/dtic/tr/fulltext/u2/a114865.pdf)
[33] LA BSS contract with FUGRO USA Ltd. April 25, 2018. (https://cityclerk.lacity.org/CouncilAgenda/AttachmentViewer.ashx?AttachmentID=69831&ItemID=67098)

building a data infrastructure to log complaints about potholes, California could pioneer a data platform that actually collects the ground-truth for regional infrastructure, paving the way for a new paradigm of municipal service delivery.

Over the years, ARGO has matured an approach of integrating open-source technology with cloud infrastructure to create data pipelines that do exactly this. SQUID or Street Quality IDentification uses Open Street Cam (https://openstreetcam.org/) to collect ride quality and street imagery to create a ground-truth about the status of local street infrastructure. The data can be parsed to construct infrastructure dashboards at a quarterly time-scale.

Consider that on a single day, a single vehicle driving at 30 mph for 5 hours, can collect ground-truth data on 150 miles of local streets. It would take fewer than 8 vehicles driving at this rate under 1 work week to collect ground-truth data on all of LA's 5,500 center-lane miles of local street infrastructure at a fraction of the cost that LA is currently spending for marginally higher quality data.

This data collection need not be limited to just pavement quality. It can extend across street signage, street furniture, property tax assessments, clean streets programs, and blight assessments.

Computer Vision methods can then be applied to these images to automatically identify features of the street such as cracks, signs, potholes, manhole covers and other useful information to support effective asset management.

The SQUID approach can also be repurposed to evaluate the quality of bike lane infrastructure and inform proactive interventions to support progressive mobility solutions. As SoCal Connected reported earlier this year in "Cycle of Disrepair,"[34] the effect of poor bike lane infrastructure has not only been fatal for bike commuters in LA but also represents a large financial burden to the city in the form of litigation from injured commuters amounting to tens of millions every year. There is a clear need for the city to be more responsive and do more with less. This can best be achieved using a data-driven approach that is currently absent at the local infrastructure level.

ARGO has the requisite experience to deploy a cost-effective data platform to realize a "ground-truthed," low-cost, anticipatory response paradigm that could potentially save Southern California cities billions while also improving service delivery in a manner that adopts open-source and transparency.

[34] SOCAL Connected. Cycle of Disrepair. (https://www.kcet.org/shows/socal-connected/cycle-of-disrepair)

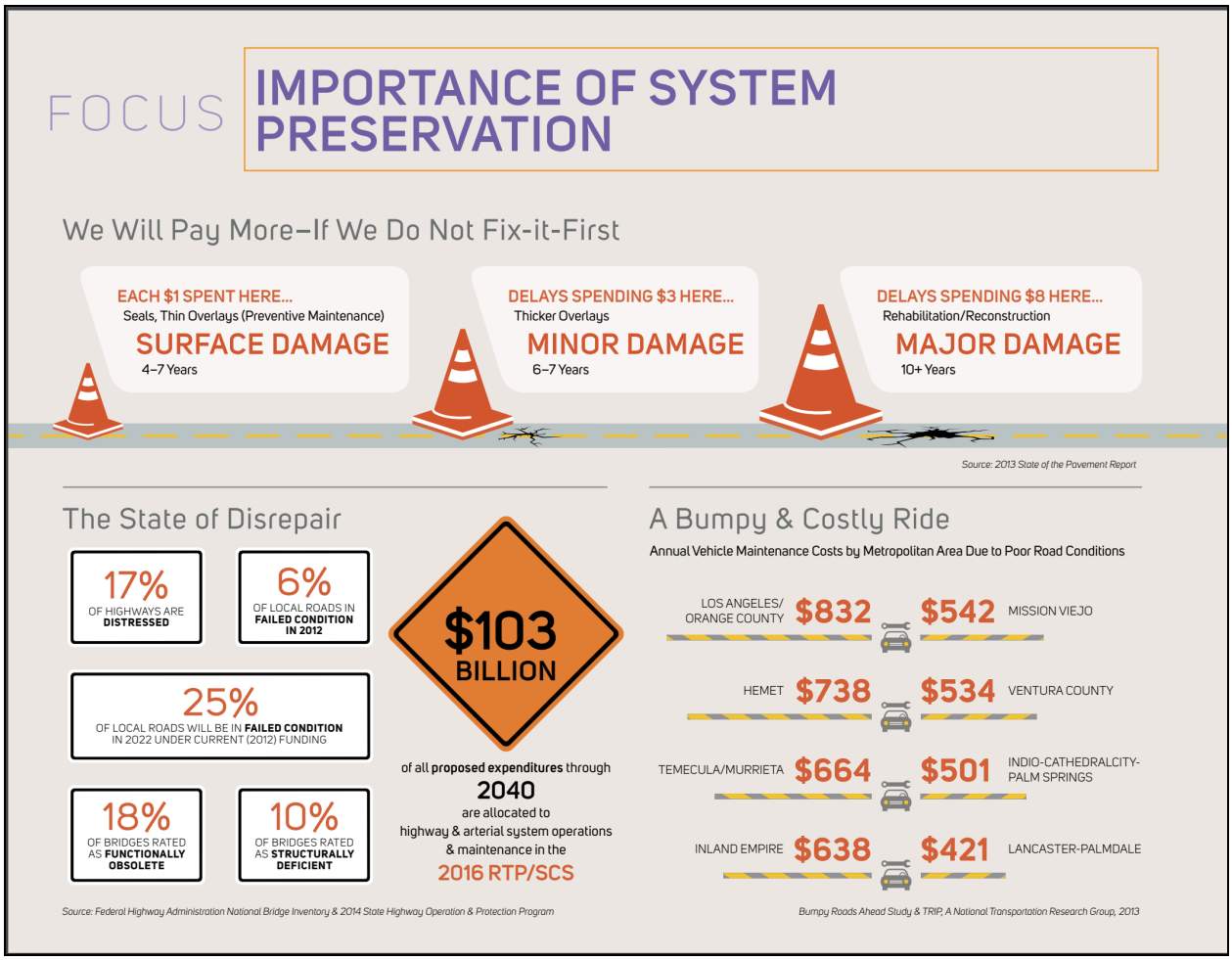


Figure 5: ARGO's SQUID approach responds to the California's street maintenance goals.

Figure 6: ARGO's computer vision model built on a training set of 100,000+ images of street quality. This model automatically detects large street defects using high-quality street imagery.

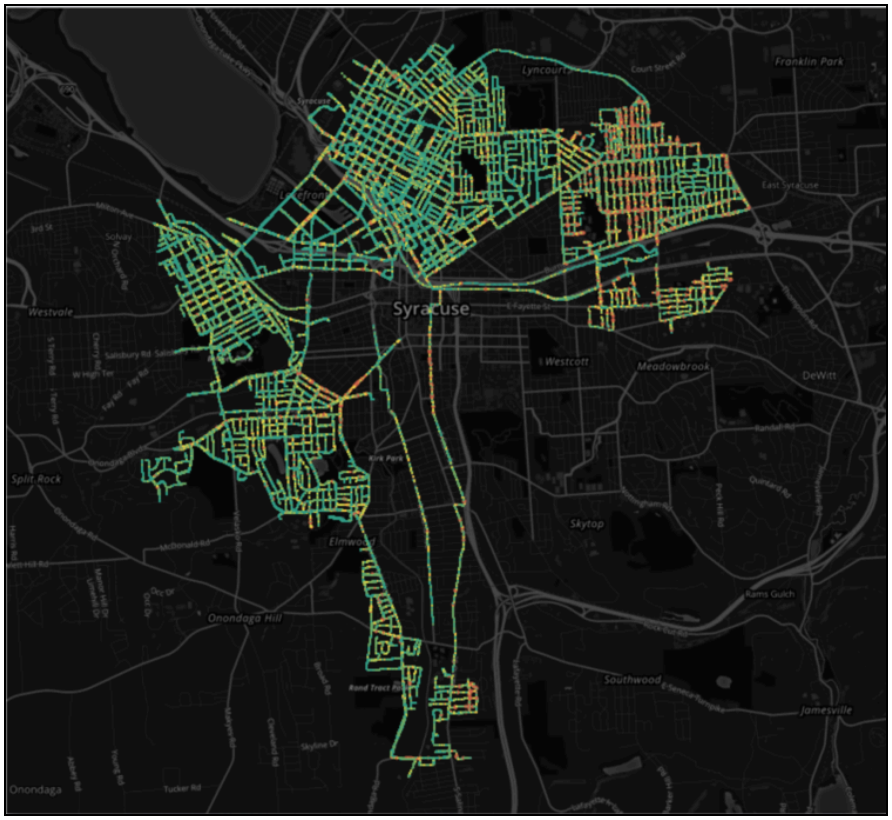


Figure 7: ARGO's Project SQUID deployed in Syracuse, NY in 2016. This map represents data collected from 110,000 images covering over 75% of Syracuse's entire street grid. The data was collected in 10 days.

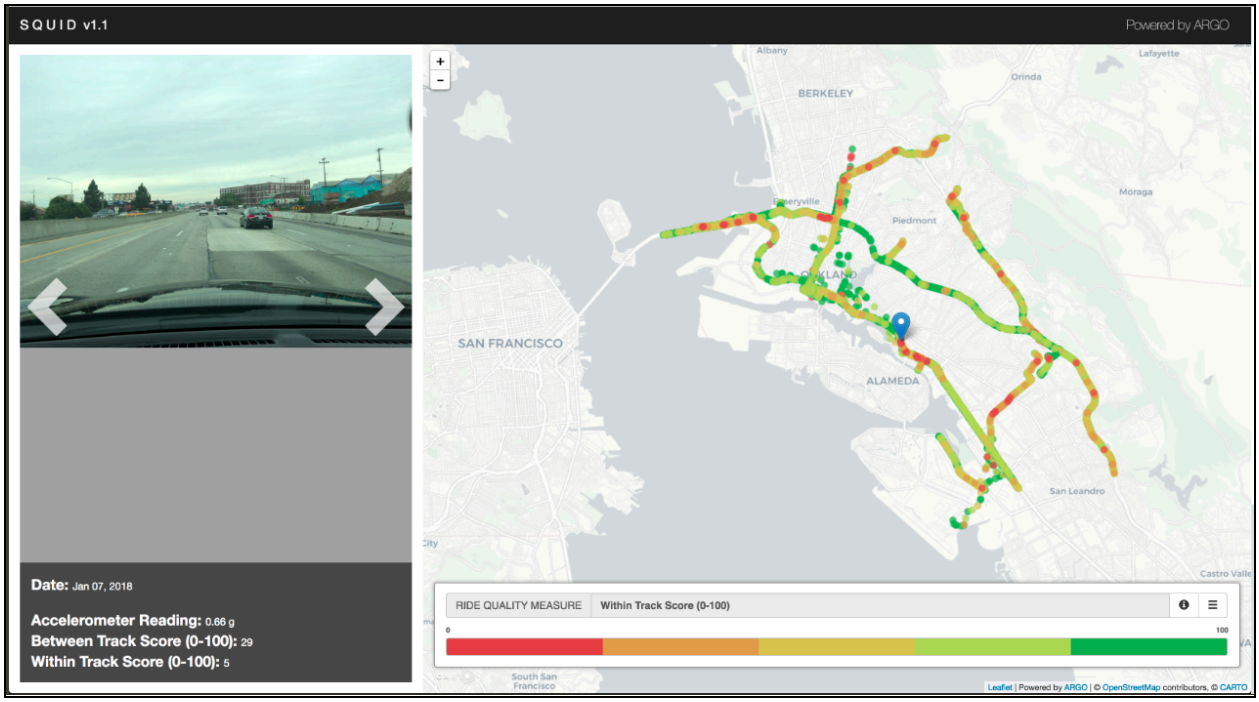


Figure 8: ARGO's custom street quality dashboard built on top of data collected by the Open Street Cam App.

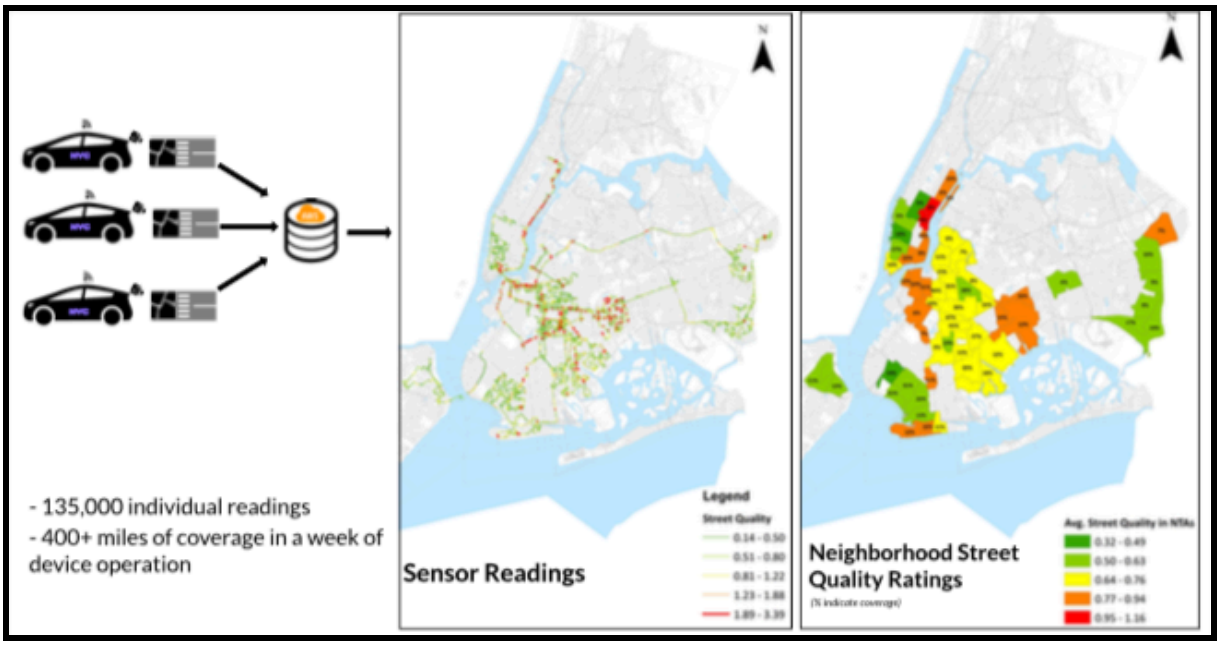


Figure 9: A single vehicle with low-cost sensors is capable of collecting 150 miles worth of ground-truthed data about local street infrastructure in a single day. Performing complete digital surveys across the entire street grid frequently can empower cities to move to a proactive response paradigm and allocate resources more effectively.

Low-cost real-time bus tracking data infrastructure

Access to real-time transit information has been linked to overall satisfaction with transit service,[35] increases in ridership,[36] and substantial increases in farebox revenue.[37] If cities could simply increase practical availability of transit information, they could realize outcomes similar to increases in transit service itself. Encouragingly, this missing layer of coordination between providers and users amounts to a conceptually simple piece of technology.

Less encouragingly, the legacy technology in this space can be exorbitantly expensive. Santa Barbara pays upwards of $35,000 per bus to provide real-time bus tracking services. While all cities and citizens could benefit from a real-time transit information system in principle, this is not an option for all cities in practice. Forward-looking municipalities have to find a new approach. ARGO's expertise can leverage the recent explosion in low-cost sensor technology and inexpensive cloud infrastructure to deploy bus tracking technology at a fraction of the cost to current market suppliers.

Consider that a $10 wifi-enabled Raspberry Pi microcomputer can use wifi-positioning technology to obtain its location with an accuracy of a few meters. This data can then be streamed in real-time to the cloud representing a solution that is orders of magnitude cheaper than legacy AVL/GPS systems. Knowing where the next bus is could be a key service offering in smaller cities to enable not only an increased reliance on public transportation leading to reduce congestion, but also ground-truthing bus operations to measure service reliability.

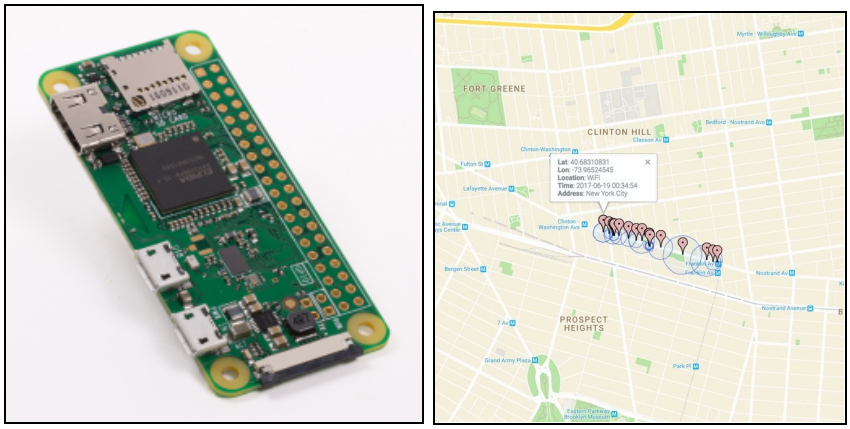


Figure 10: (L) A $10 Wi-fi capable Raspberry Pi Zero.
(R) Testing wifi positioning system on a bus.

Planning for Autonomous Futures

[35] The real benefits of real-time transit data – Sidewalk Labs. (https://medium.com/sidewalk-talk/the-real-benefits-of-real-time-transit-data-1fee19988b73)
[36] Science Direct, Ridership effects of real-time bus information system: A case study in the City of Chicago. (https://www.sciencedirect.com/science/article/pii/S0968090X12000022)
[37] Science Direct, The impact of real-time information on bus ridership in New York City. (https://www.sciencedirect.com/science/article/pii/S0968090X15000297)

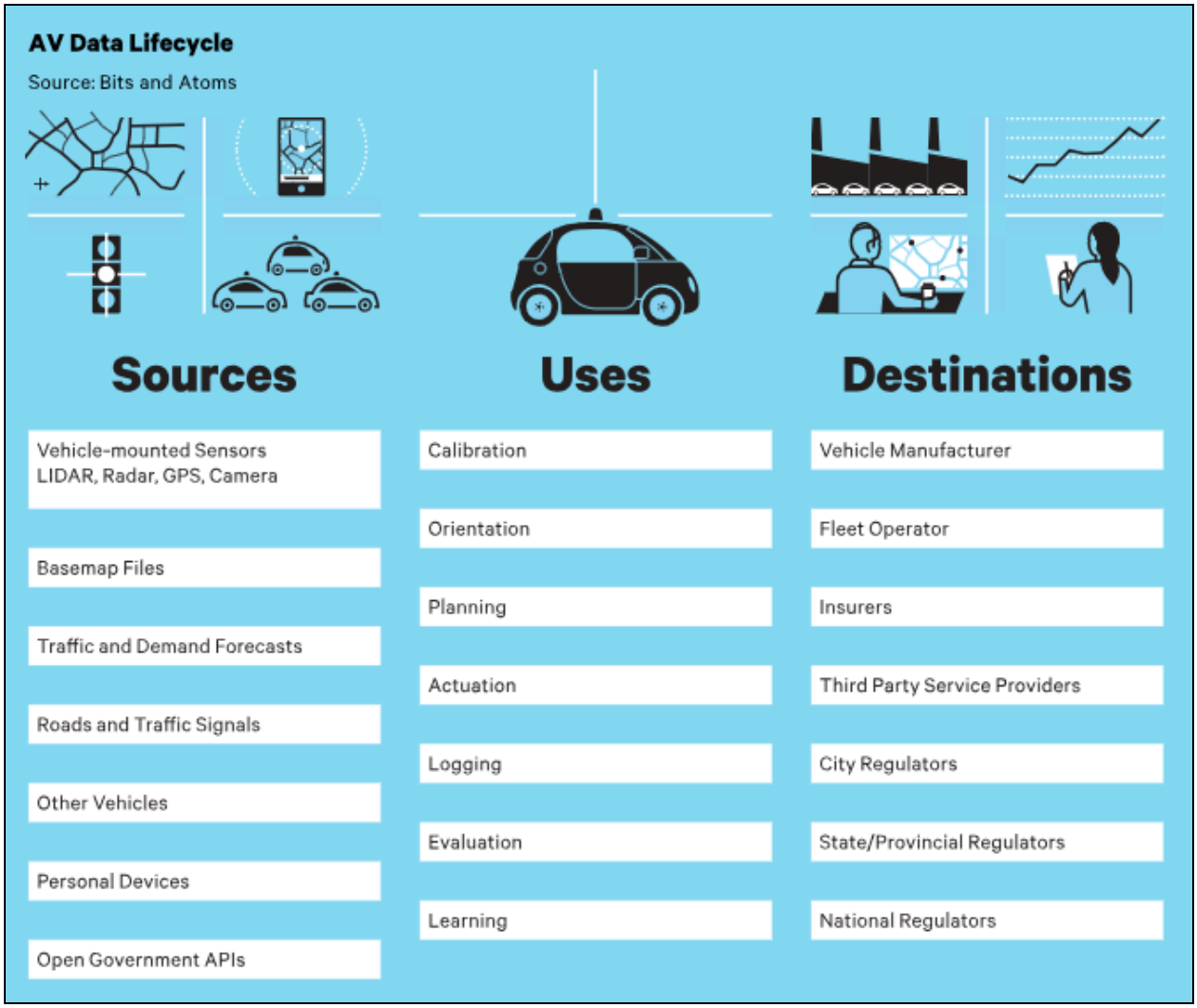


Figure 11: Deploying autonomous vehicles at scale will require a comprehensive & robust data infrastructure followed by unprecedented collaboration around data from the public and private sectors. *Source: Autonomous vehicles: Future Scenarios National League of Cities Center for City Solutions (http://avfutures.nlc.org/)*

"Transportation planners and engineers must be ready to anticipate how shared mobility and connected and automated vehicles will revolutionize mobility - creating incredible opportunity to impact our environment, economic development and equity in positive ways. However, if ignored, these technologies can have a devastating impact that could induce greater sprawl, inequity and other challenges for the city and region."[38]

While buses and potholes represent the past and present challenges of regional infrastructure, autonomous vehicles represent the future and they are currently on California's streets. As of

[38] Ashley Z. Hand. Urban MOBILITY in a Digital Age (http://www.urbanmobilityla.com/strategy/)

this submission, Waymo, an Alphabet company, has begun an aggressive expansion plan that includes purchasing up to 62,000 vehicles and eyeing deployments in various regions of California.[39]

Vehicle autonomy thus far has been a private, luxury service but represents an immense opportunity to serve various public mobility needs, at scale. Consider that social workers and visiting nurses can be transported to their client's homes without experiencing the fatigue of maneuvering traffic or changing routes. Consider that the elderly and disabled can engage with the outside world more frequently without needing to rely on another person.

Data about streets has not changed in decades. Different cities use different formats to store and publish data. In many cases, this data is not even machine-readable and results in a painstaking, manual, and expensive process of digitizing local street data. ARGO has a good professional relationship with the technical team and leadership at SharedStreets, a Bloomberg Philanthropies funded initiative enabling interoperability of urban mobility data. The initiative describes itself as follows.

"SharedStreets is a non-profit digital commons for the street. It's a data standard and a platform that serves as a launching pad for public-private collaboration and a clearinghouse for data exchange. SharedStreets is a system for sharing actionable information across city and company lines, and helps communities better understand and manage their streets."[40]

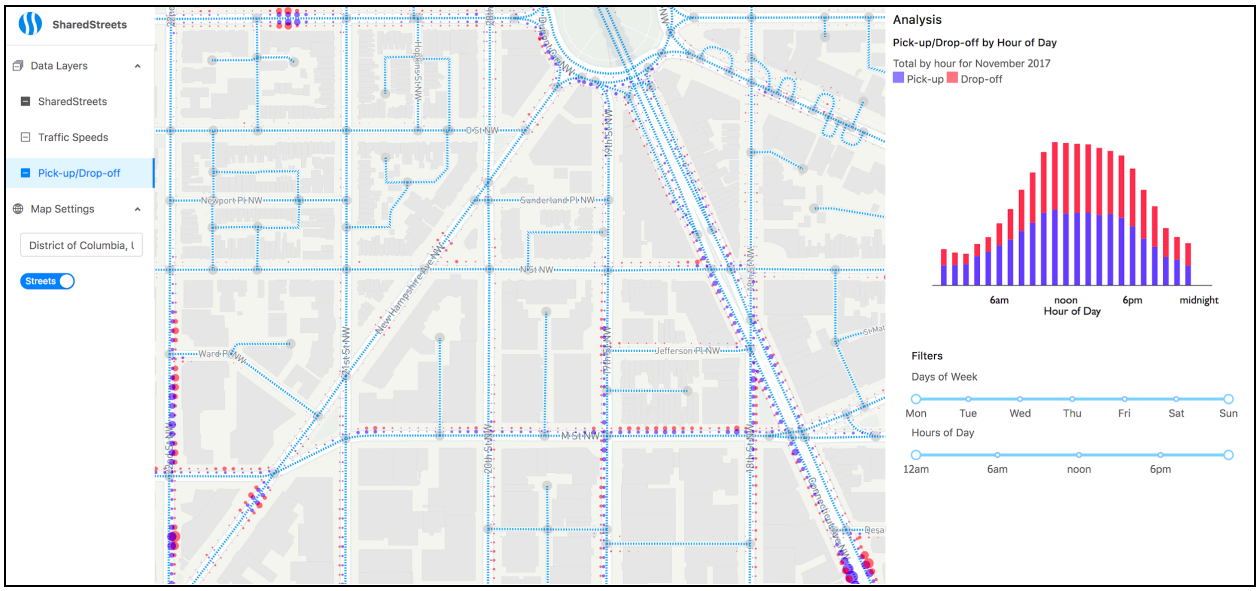


Figure 12: Taxi pick-up and drop-off volume by time of day near Dupont Circle,

[39] Quartz.com There are about to be Waymo cars on the road. ((https://qz.com/1294001/waymo-ordered-62000-chrysler-pacifica-minivans-from-fca-as-it-rolls-out-autonomous-ride-hailing-taxis-across-the-us/)

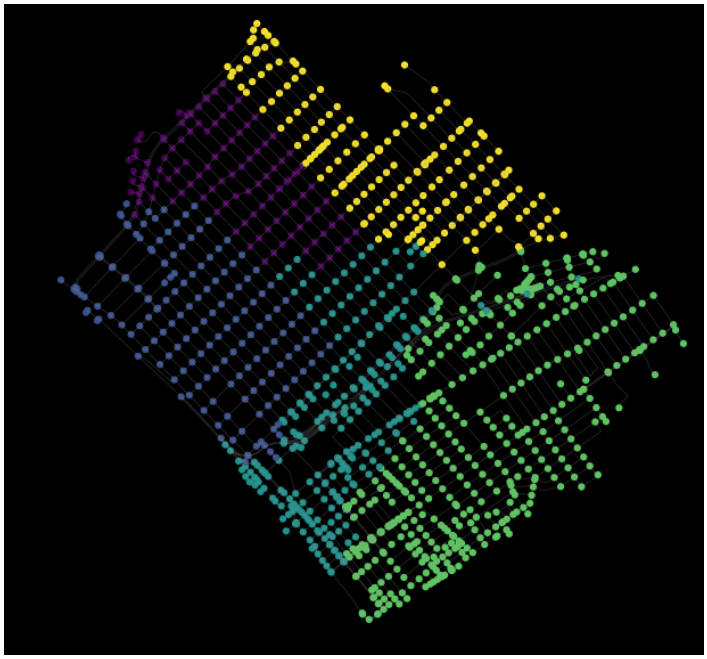
[40](www.SharedStreets.io)

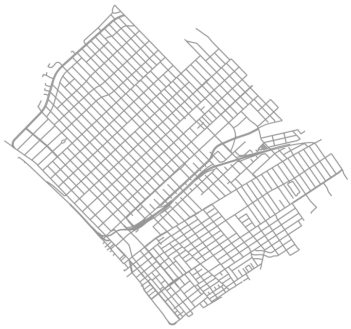
Washington, DC. Data from DC Department of For-Hire Vehicles. (Source: Shared Streets)

To realize the idealized outcomes of vehicle autonomy in an equitable manner, AV manufacturers need to ensure that their vehicles are able to route across every street in a city. Today this remains to be seen and largely unrealized because of factors ranging from fragmented street data, poor street maintenance - where an absence of lane markings make it challenging for an AV to navigate, to maintaining expensive and proprietary mapping data infrastructure.

ARGO has already begun developing route modeling tools by operationalizing algorithms such as the Postman or Route Inspection problem. Algorithms to create automated route planning for all streets in a city is considered to be a computationally complex problem[41] but with the assistance of inexpensive computing in the cloud and imaginative approaches to parallelize the computation, we believe that it is possible to develop a general purpose route planning tool for a public service autonomous vehicle to automatically route across each and every street in a given city. In the near term, such an algorithm can go a long way to improve resource allocation for current municipal operations such as street cleaning and digital street surveys.

[41] Chen Y., Polack F., Cowling P., Remde S. (2017) A Comparison of One-Pass and Bi-directional Approaches Applied to Large-Scale Road Inspection. (https://link.springer.com/chapter/10.1007/978-3-319-53982-9_11)





We can easily see that Santa Monica has a more orthogonal, grid-like street network than Moraga does. This is clearly reflected in its polar plot:

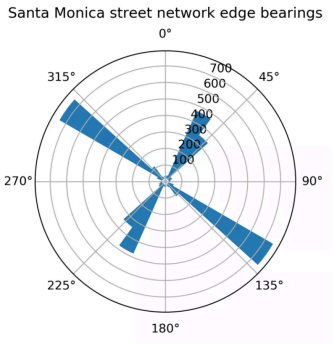


Figure 13: (L) Fluid Communities is an algorithm to automatically categories network graphs into distinct communities.[42] (R) The geographic orientation of street network is yet another method to categorize street segments into distinct groups.[43] Both techniques are helpful for complete route planning algorithms.

High-quality mobility and ride-sharing data from companies like Waymo, Uber, Lyft, or Tesla can be made available to public agency partners and the extended research community to collaboratively develop strategies with Streets Data Collaborative members to provide mobility solutions targeting populations who most need them.

2.3 Potential Case: Building a Housing Data Collaborative

BLUF: California's housing crisis is compounded by the incredible cost and complexity of building here. This memorandum proposes to improve the implementation of the state's existing, legislatively enacted

[42] Fluid Communities: A community detection algorithm. (https://arxiv.org/abs/1703.09307)
[43] GeoffBoeing.com. OSMnx. Street Network Orientation. (http://geoffboeing.com/2018/02/street-network-orientation/)

goals (AB 1483, Grayson 2019) to make housing development costs and approval procedures more transparent through 1) a philanthropically seeded housing data collaborative to accelerate adoption of improved housing data standards and 2) make the updated data requirements subject to a public records act request.

The Problem in Brief

"When developers can't even find an estimate of what their fees will be, what affordability requirements they must meet, or what impacts they will be required to mitigate for, they are likely to pull up stakes and move to a friendlier jurisdiction (or state for that matter)."

-Assemblymember Tim Grayson

California's housing crisis is the number one factor making the state unaffordable to live. In most of the state, no one making a typical income can afford a typically priced home.[44] This cost of living prices out more and more Californians, exacerbating inequities and the legacy of practices like redlining that have led to a massive racial wealth gap. The price of housing hurts California's ability to attract talent and invest in the future.

The cost and complexity of building new housing in California exacerbates this crisis. As well articulated by Assemblymember Grayson, too often what the requirements and fees a new housing development is subject to are unclear. Uncertainty kills new housing. AB 1483 was designed to address the underlying lack of transparency in the housing permit process. Yet by the latest count, less than a quarter of jurisdictions actually complied with the laws' requirements[45]. Lack of local technical capacity and the absence of state enforcement mechanisms have led to poor adoption.

Proposed Solution

ARGO proposes to accelerate AB 1483 implementation through a philanthropically funded data collaborative to align on new data standards and include the requirements for local municipalities in AB 1483 as part of the California Public Records Act.

The new housing data collaborative would build from the existing AB 1483 workgroup by providing a focused team with civic data science skills to develop and iteratively refine data standards. The Mobility Data Specification provides a great example of the type of collaborative approach to developing a new standard and also supporting software tooling around the new specification.

That technical support for local municipalities would be balanced with enhanced enforcement. The existing law has no penalties for non compliant cities. This memorandum proposes to include the AB 1483 local government data requirements as part of the California Public Records Act. That was successfully utilized by Senator Hertzberg's SB 272 local government open data inventory law.[46] That amendment should be phased to follow the development of machine readable data collection protocols and standards developed by California's Housing Data Collaborative.

[44] CA FWD, "Building Racial and Economic Equity through Home Ownership."
[45] SPUR, "How Much Does It Cost to Permit a House?"
[46] The EFF crowdsourced a repository of SB 272 data inventories, which saw much greater compliance.

Standardizing cost, zoning, and permitting data will form the foundation for streamlined housing production. Increased supply of housing has a direct impact on prices for both renting and owning a home,[47] benefiting California residents broadly. In addition, streamlining permitting through data standards would benefit in particular smaller developers that lack the capacity for compliance in a complex and opaque regulatory environment. Those firms are more likely to be owned by Californians of color.

Effectively implementing AB 1483 also provides the foundation for a generational shift in how California plans for and permits urban growth.

Vision for the future

How we plan as a state and process proposed developments has not fundamentally changed since the seventies. Common sense streamlined data access also forms the foundation for new applications and approaches to planning for and permitting new housing, as well as other infrastructure and urban development. The housing data collaborative could build from its successes in accelerating data standard adoption to iteratively pilot new protocols for analyzing project benefits, impacts, and means of engaging the public. Pilots of existing augmented reality technology to showcase a next generation alternative approach to urban planning and development could be implemented today.

[48]

Resources:

- HCD Data Strategy: https://www.hcd.ca.gov/docs/data-strategy.pdf

[47] Noah Smith, "I have seen the future of housing."
[48] Images courtesy of InCitu.

- Bill author thread: https://twitter.com/AsmGrayson/status/1418646918150856704

Section 3: ARGO's Kraken Public Data Platform

ARGO's Kraken data integration platform currently automates data integration from a coalition of California water utilities serving 23 million people. Building that battle-tested platform has yielded many pragmatic insights that will benefit the development of the RDP. It also highlights ARGO's ability to complete a technically challenging data infrastructure project while navigating an environment of high uncertainty involving numerous stakeholders and a highly sensitive political landscape.

ARGO also works continuously to improve our data infrastructure practices and utilizes a simple internal application to track the inevitable errors that arise in ingesting public data. This tool templatizes root cause analyses and logs standardized classes of incidents. This allows any member of our team to follow appropriate next steps from a library of turn-key solutions as opposed to reacting to each incident as an ad hoc fire.

Public and particularly administrative data inevitably has data quality issues arising from practices like changes in classification systems and simple human error. For example, a public report of urban water demand can occasionally see changes in naming conventions which can break an automated data parser. Addressing that type of data categorization error requires flexible tooling to first identify the problem and adaptive internal practices to ensure smooth operation of ARGO's data infrastructure.

ARGO operates under the belief that data standardization must proceed from deep domain expertise and understanding of the underlying public challenge. Data integration for the sake of data integration serves no public purpose. ARGO thus standardizes datasets through close collaboration with our local government partners in a leading "data collaborative" model described in greater detail in Section 2. That model involves stakeholders throughout the process including prioritizing the initial datasets to be integrated.

Further, that data collaborative model segments technical staff working with data from decision-makers who will utilize the analytical results of that work. For many technical tasks, it is important to have buy-in from leadership and then empower the staff working day to day with data to have the space to deliver the work.

Using this data collaborative model, ARGO would then establish a transfer protocol to obtain regular updates for key data sources. This is the first stage of the Kraken data platform, where data is transferred to a central location, at first either manually or programmatically. From this central location, an extract, transform and load (ETL) process is then used to standardize data

from diverse sources and pipe it to the appropriate location, whether that is a public agency website, GIS portal or backend geodatabase.

Team ARGO has successfully deployed enterprise-grade ETL processes across our CaDC community of California urban water utilities. These include experience integrating land use data across the state from our work on Governor Brown's framework to "make water conservation a California way of life." In addition, ARGO has a robust protocol for standardizing metered water usage data across local water utilities and we believe that there are some synergies between those protocols and the broader need for data standardization (in for example parcel level land use or municipal financial audits). One aspect of ARGO's approach involves developing a mapping between specific local utility customer categories and a standardized set of customer categories across California. This standardization enables comparisons and analysis across jurisdictions.[49] A similar approach could be applied to land use data. As another example, ARGO has integrated structured water meter data with unstructured water rate data, often contained in PDF documents by leveraging our Open Water Rate Specification which is, to the best of our knowledge, the first open source data specification for water rates.[50]

```yaml
---
metadata:
  ...
rate_structure:
  RESIDENTIAL_SINGLE:
    service_charge:
      depends_on: meter_size
      values:
         3/4": 14.65
         1"  : 16.77
         2"  : 25.83
    tier_starts:
      - 0
      - 15
      - 41
      - 149
    tier_prices:
      - 2.87
      - 4.29
      - 6.44
      - 10.07
    commodity_charge: Tiered
    bill: commodity_charge+service_charge
```

Figure 14: An example of ARGO's Open Water Rate Specification. (OWRS)

ARGO's Kraken infrastructure attempts to automate as many of these steps as possible after an initial consultation. Data collection scripts are scheduled on a periodic basis, and quality

[49] See here for the California Data Collaborative onboarding protocol: (https://docs.google.com/document/d/1ppWv3T2Sjr0zNP40PB5C_4NAREC1n_JtOvudpvZBNmE/edit?usp=drive_web&ouid=108248130259539340429).
[50] ARGO Open Water Rate Specification. (https://github.com/California-Data-Collaborative/Open-Water-Rate-Specification)

checks are written as programmatic constraints. This approach requires an upfront investment when initially ingesting data from a new source, but it saves staff time in the long run when data updates occur frequently. In cases where fully automated data integration is not practical or desirable, data can be loaded to a temporary staging location for manual review in alignment with current data policies. For example, while some quality checks on spatial data are easy to automate, others may be best handled through a manual review and approval process. Kraken is highly flexible and can provide partner government staff automated notifications via email or other communication channels upon request.

The Kraken platform is cloud-based and built on top of the industry standard cloud services provider Amazon Web Services (AWS). Data transfers and archiving are done using Amazon Simple Storage Service (S3) which provides low cost, highly scalable and extremely reliable object storage. In the Kraken platform, S3 serves as a secure "data lake" that holds raw unprocessed datasets. These data are preserved in their original form, forming a read-only archive of past data transfers. These original data are then read and processed in the next step of the pipeline.

The ETL process is performed using the open-source Apache Airflow workflow software to define directed workflow graphs of processing steps. Airflow has become the de facto standard workflow software for data engineering among many software companies--including Airbnb, where it was originally developed--because of its flexibility and ease of use.[51] Processing steps (called "tasks") are arranged into workflows that define dependencies among different steps in the pipeline. The workflow for ingesting any particular data source may include steps such as standardizing attribute names, filtering invalid data and enriching primary datasets with useful secondary data sources for analysis.

After processing, the cleaned and enriched data can be loaded wherever it is needed. For some geospatial datasets, this might mean a direct upload to Data.CA.gov or another open data portal. For others, it might make sense to load the clean data into an integrated data warehouse where it can power web services and tools provided to local jurisdictions, or be exposed to the public using an API. Note that ARGO has deep experience developing data integration from legacy IT systems such as local water utility billing systems and water conservation rebate data at both the local and regional levels.

All of this ETL is possible using Airflow, which uses Python to define workflows but otherwise gives the data engineer freedom to define their own data processing steps. A key advantage of using open source workflow automation platforms such as Apache Airflow is that it can be repurposed to develop automated testing frameworks to ensure:

[51] The official GitHub provides an impressive list of Airflow users. (https://github.com/apache/incubator-airflow/blob/master/README.md). The canonical O'Reilly technology media series also devotes an entire chapter to Airflow in "Agile Data Science 2.0." (http://shop.oreilly.com/product/0636920051619.do)

- Overall data quality i.e. sanity checking structured data for errors in schema or values.
- Application testing i.e. Checking application integrity under different scenarios.
- Stress and Load Testing: Testing ARGO's public data infrastructure against capacity-related metrics.
- Smoke testing to ensure that key applications are available and stable at all times. Examples include testing the file ingestion component of the data platform where every morning a dummy file will be uploaded, parsed, and visualized through the data platform.

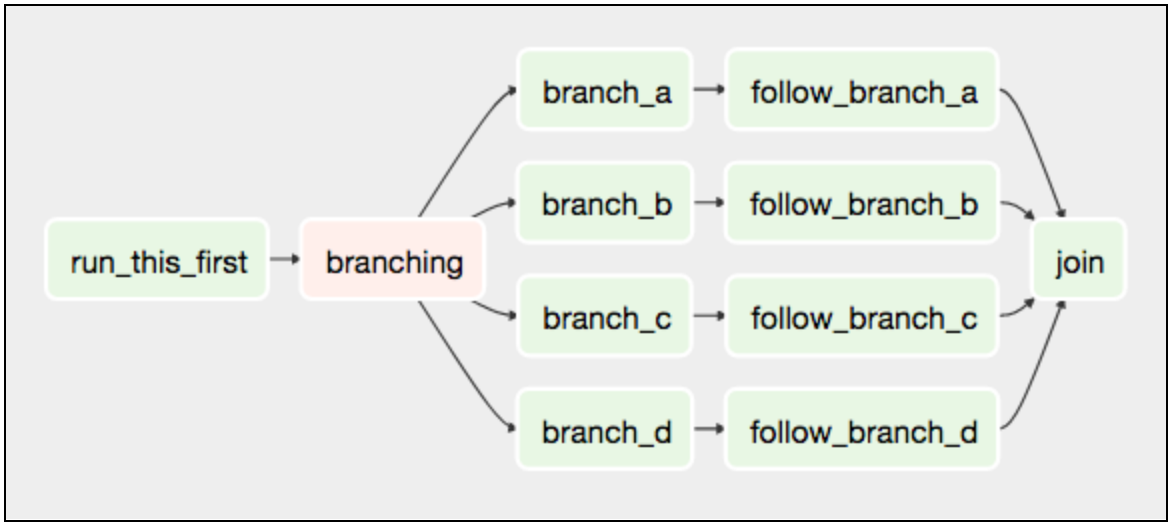


Figure 15: A canonical workflow in Apache Airflow. This system can be not only employed for core data engineering tasks but also govern an enterprise testing framework.

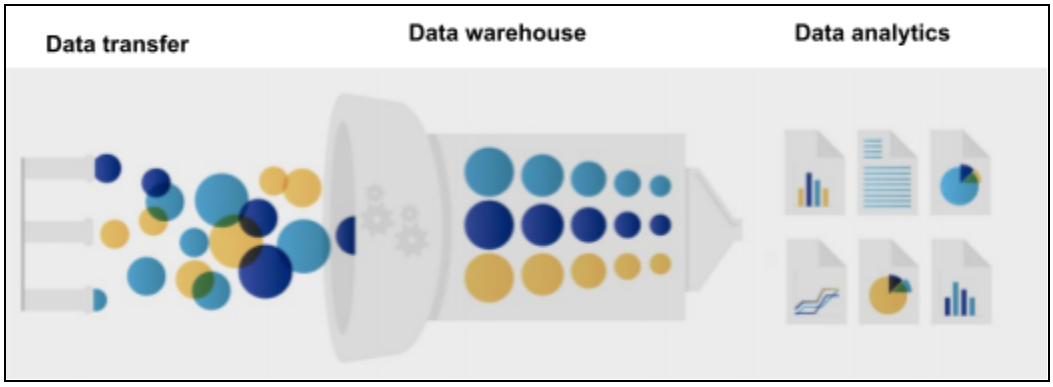


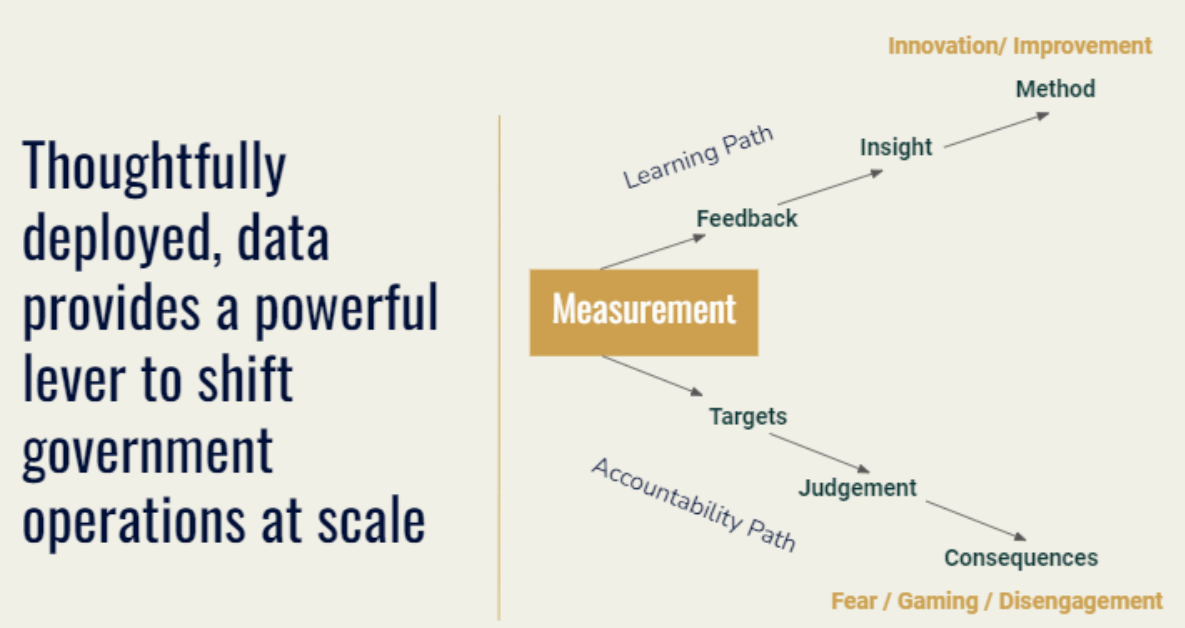
Figure 16: Video walkthrough of Kraken: (https://www.youtube.com/watch?v=MW844gMNr2U).

Data usage agreements will be crafted to address the privacy or sharing concerns of data platform stakeholders. ARGO will draw on our experience with the California Data Collaborative

where we currently process data transfers from water utilities across the state using a streamlined data sharing procedure and a standard non-disclosure agreement.[52]

Towards measuring public outcomes by default

Robust data infrastructure is an important enabler of quality measurement. Data collaboratives provide a useful model to support public managers on their learning journey to implementing adaptive, digitally native practices.



Credit: Ken Miller.

Ultimately the goal is to nurture an ecosystem. Here there is ample opportunity to experiment with bounties and novel market mechanics to open up the very insider-ish state and local government consultant industrial complex.

Similar to how dune dot xyz provides a robust marketplace for web3 analysis but for computation social science. That would be a way to potentially close the massive evidence gap and make rigorously measuring the impact of public investment the new normal. Interfluidity has an excellent vision for what we might build:

"Ultimately, we should want to generate a reusable, distributed, permanent, and ever-expanding web of science, including conjectures, verifications, modifications, and refutations, and reanalyses as new data arrives. Social science should become a reified public commons. It should be possible to build new analyses from any stage of old work, by recruiting

[52] ARGO's approach to data sharing is described in the following blog post.(http://californiadatacollaborative.org/blog/2016/6/15/lowering-the-transaction-costs-of-sharing-californias-water-data)

raw data into new projects, by running alternative models on already cleaned-up or normalized data tables, by using an old model's estimates to generate inputs to simulations or new analyses."

Background on key Argonauts

Accelerating Reform in Government Operations ("ARGO") operates as the world's first public data utility. Similar to water or electricity, ARGO believes public data is a public resource that deserves technocratic excellence and needs to be stewarded for the public good.

Uniquely, ARGO isn't yet another tech company pitching software products but rather has a demonstrated track record of helping local governments build capacity and leverage civic data science talent through a unique shared service model. ARGO operates as a fiscally sponsored project through the National Center for Civic Innovation, a 501(c)3 nonprofit incubator.

ARGO's data scientists have worked in-the-trenches in New York University's Administrative Data Research Facility, led urban technology initiatives at New York's New Lab hardware technology incubator, and delivered automated data parsing pipelines for a $50 million startup (Enigma Technologies) that powers the world's most comprehensive repository of public data.

ARGO is headquartered out of the Los Angeles Clean Technology Incubator ("LACI") located in the Arts District. ARGO can be reached by mail at ATTN: ARGO 525 S Hewitt Street Los Angeles, CA 90013 and by calling 213.358.6500 and asking for ARGO staff.

Key members of ARGO's team

Patrick Atwater envisioned, developed and collaboratively launched ARGO's big water data project in partnership with leading local water managers. He has worked as a consulting data engineer for a multi-million dollar venture-backed startup, ran the numbers for the State Water Contractors on Governor Brown's Delta Fix project and co-authored the feature June 2015 American Water Works Association article on how data science can help water utilities adapt to climate change. As part of Team ARGO, his civic data science work has been featured by the Knight Foundation and Fast Coexist. He studied mathematics and PPE at Claremont McKenna College, holds a master's degree in urban informatics from NYU CUSP, and completed the Coro Fellows program. He currently serves on the board of the Los Angeles Education Partnership and has written extensively on California, technology and the future of government operations.

As a Coro Fellow, Patrick Atwater spent his government placement at SCAG and explored opportunities to modernize how SCAG provided comments per the California Environmental Quality Act using existing digital tools. In addition, Patrick spent 1.75 years as an analyst at

Public Financial Management ("PFM").

Christopher Tull is a data system expert with a focus on delivering customized data storage and analytic solutions to support local governments. He has served as lead data scientist with the California Data Collaborative where he designed and built data infrastructure to process, store, and analyze metered water use data from water utilities across California, including both traditional monthly billing and hourly AMI data. Prior to that, he worked in the Urban Intelligence Lab at New York University where he developed the official energy benchmarking analysis website for the City of New York and forecast energy use for all 1.1 million buildings in NYC. His software powers ongoing initiatives in the public, private and nonprofit sectors as well as in academia where his work has been published in peer-reviewed journals and top industry conferences. He holds an M.S. in Urban Informatics from New York University and a B.S. in Computer Science and Mathematics from CSU Channel Islands.

David Marulli is a data scientist working across the disciplines of data engineering, analysis, and design. With the California Data Collaborative, he has led the development of the Statewide Efficiency Explorer tool that delivered the world's first estimates of Governor Brown's water efficiency framework for 5% of the $3 million the state budgeted for the project. He is leading other ongoing statewide data integration efforts including designing the architecture of ARGO's new Streets Data Collaborative. David is also expanding the Efficiency Explorer to the create a common operational picture of water reliability that includes supply and future system stress. Prior to working with ARGO, his work was sponsored by NYU's Rudin Center for Transportation Policy and Management, and the Urban Intelligence Lab. He holds a Masters in Urban Informatics from NYU and Bachelor degrees in Physics and Philosophy from UW-Madison.

Varun Adibhatla has over 12 years experience developing and deploying enterprise-grade data technology at the intersection of industry, academia, and government. He has built, operated, and maintained data infrastructure in Wall Street, specifically for High-Frequency Global Foreign Exchange Trading systems.

Varun has also served the National Resource Network, a federally funded initiative to support economically challenged cities in implementing policy and technology solutions. He led research on NRN's foundational report on economically challenged cities in the report California Hidden in Plain Sight.[53]

[53] NRN Report: (http://www.nationalresourcenetwork.org/en/Document/306226/Hidden_in_Plain_Sight_Why_Californias_Economically_Challenged_Cities_Matter).

Varun served as the Director of the Urban Tech Hub at New Lab, a unique ecosystem located in the historic Brooklyn Navy Yard where he bootstrapped and launched the multi-million dollar Urban Tech Hub RFP with the NYC Economic Development Corporation.[54]

Lastly, Varun is an experienced cloud data infrastructure specialist with skills in building data organizations that focus on agility and reliability. He has also worked as a Site Reliability Engineer for an online video streaming business with a global audience of millions.

Graham Henke is data systems engineer with a particular focus on automation and user experience. As part of ARGO's team powering the California Data Collaborative ("CaDC") coalition of water utilities, he designed and built data infrastructure to automate the ingestion and storage of water use data. At NYU CUSP, Graham worked on a number of projects where his responsibilities included project management, user experience, and technical training related to data platform known as the Administrative Data Research Facility, which serves as a data sharing and analysis platform for local, state, and federal government agencies. Prior to his experience with the CaDC and CUSP, Graham worked in data center operations with Apple Inc. He has an M.S. in Urban Informatics from New York University and a B.S. in Computer Science from Purdue University.

Select clips of ARGO's work in the news

- **The Obama White House**, Commitments to Action on Building a Sustainable Water Future. (https://obamawhitehouse.archives.gov/sites/whitehouse.gov/files/documents/White_House_Water_Summit_commitments_report_032216_v3_0.pdf)
- **California Senate Hearing** for SB272 on Modernizing Government. (https://www.youtube.com/watch?v=ToKLiXQYthg&feature=youtu.be&t=50m56s)
- A New Water Data Initiative is Changing Water Management. **KQED** (https://ww2.kqed.org/science/2016/05/16/a-new-data-initiative-is-changing-water-management/)
- Land of Sod. **Harper's Magazine** (https://harpers.org/archive/2016/09/land-of-sod/)
- Smart Cities: The Pothole Problem. **FOX News** (https://www.youtube.com/watch?v=QzMCTIdxpoc&feature=youtu.be)
- Why are the streets always under construction? **New York Times** (https://www.nytimes.com/interactive/2016/08/18/nyregion/new-york-101-streets-repair-and-maintenance.html?_r=0)

[54] Article on Varun's work at NYC New Lab. (https://technical.ly/brooklyn/2016/09/01/varun-adibhatla/)

- A new cheap way to quickly map your city's potholes. **Fast Company** (https://www.fastcompany.com/3059201/a-new-cheap-way-to-quickly-map-your-citys-potholed-streets)
- A 21st-century proposal for Street maintenance. **Staten Island Live** (https://www.silive.com/news/index.ssf/2016/04/a_21st_century_proposal_to_map_potholes_repair_streets_efficiently.html)
- SQUID to give Syracuse better feel for needed road repairs. **WRVO** (http://wrvo.org/post/squid-give-syracuse-better-feel-needed-road-repairs)
- Filling potholes faster. **Metro New York** (https://www.metro.us/new-york/local-innovators-aim-for-access-to-city-contracts-through-new-firm/zsJohE---bWgSVgQYksuY)

“Everything is possible until it’s not possible.”
-Varun Adibhatla

ARGO
Applied Research in
Government Operations