

DATA ECOSYSTEM PROJECT FINAL REPORT & RECOMMENDATIONS

bcANALYTICS
MAY 2016

PREPARED BY

The buildingcommunityWORKSHOP ([bc]) is a Dallas based nonprofit community design center seeking to improve the livability and viability of communities through the practice of thoughtful design and making. We enrich the lives of citizens by bringing design thinking to areas of our city where resources are most scarce. To do so, [bc] recognizes that it must first understand the social, economic, and environmental issues facing a community before beginning work.

SUPPORTED BY

As the largest community foundation in Texas and one of the largest in the nation, Communities Foundation of Texas (CFT) works with families, companies and nonprofits to strengthen our community through a variety of charitable funds and strategic grantmaking initiatives. The foundation professionally manages more than 900 charitable funds and has awarded more than \$1.3 billion in grants since its founding in 1953. Increasing financial stability of working families is one of the two key focus areas of CFT's community impact funds. To support this area, CFT has launched the Data Driven Decision-Making (D3) Institute. The D3 Institute is designed to provide organizations that offer programs and services for low-income working families the power to accelerate their development of enduring solutions to the social and economic problems facing this population. www.cftexas.org/D3

ABOUT bcANALYTICS

Overcoming the major economic and social issues that affect the working poor requires coordinated data gathering and analysis. To address these issues, buildingcommunityWORKSHOP has developed a data and analytics team, bcANALYTICS, focused on providing high quality services to nonprofit and community based organizations. Additionally, bcANALYTICS, has begun researching and discussing a variety of ways to improve data collection, analysis, and sharing in the region.

D3 Analytics

Through Communities Foundation of Texas' Data-Driven Decision-Making Institute (D3), bcANALYTICS has offered graduating organizations an analytics package to answer pressing organizational questions - from appropriate locations for a new facility to demographic analysis of an organization's existing service area. To date, bcANALYTICS has completed reports for 10 organizations working across North Texas.

Data Ecosystem Project

In 2014, bcANALYTICS began researching national best practices in community-focused data analysis, collection, and sharing in order to identify scalable solutions for Dallas. This work has been focused in three primary areas: a review of organizations nationwide, a survey of Dallas nonprofit data community, and conversations with stakeholders working across North Texas. The project seeks to identify practical approaches to improving the ecosystem of data and information in the region.

Fee-For-Service Analytics

bcANALYTICS offers services to nonprofit and community based organizations. Services offered include data collection, analysis, cartography, and visualization. Additionally, bcANALYTICS provides support to projects housed in other units of buildingcommunityWORKSHOP.

ACKNOWLEDGEMENTS

We would like to thank the following individuals and organizations for their participation, feedback, and insight during the Focus Groups and Labs:

Jenny Eyer (*Children at Risk*), Katie McConnell (*Children At Risk*), Corron Sanders (*Children's Health*), Summer Collins (*Children's Health*), Sweetie Baidhya (*Children's Health*), Casey Gardner (*City of Dallas*), Girish Ramachadran (*City of Dallas*), Landon Gardner (*City of Dallas*), Thomas Brindle (*City of Dallas*), Tim Glass (*City of Dallas*), William Finch (*City of Dallas*), Ashwina Kirplani (*Commit!*), Robert Munding (*Commit!*), Ashley McIver (*Communities Foundation of Texas*), Sarah Cotton Nelson (*Communities Foundation of Texas*), Wende Burton (*Communities Foundation of Texas*), Benjamin Magill (*Dallas County Community College District*), Trey Bowles (*Dallas Entrepreneur Center*), Jennifer Sanders (*Dallas Innovation Alliance*), Mark Draz (*Dallas Public Library*), Duane Dankesreiter (*Dallas Regional Chamber*), Eric Griffin (*Dallas Regional Chamber*), Gary Hufstedler (*DART*), Kristin Kuhne (*Educate Texas / CFT*), Jane Massey (*Habitat for Humanity*), Richard Amory (*North Texas Food Bank*), Sue Pickens (*Parkland Health & Hospital System*), Annie Wright (*SMU*), Dylan Farmer (*SMU*), Ashley Brundage (*United Way of Metropolitan Dallas*), Galen Smith (*United Way of Metropolitan Dallas*), Anthony Galvan (*UTD*), Timothy Bray (*UTD*), Richard Perez (*Workforce Solutions*), Blake Burris, Errika Flood-Moultrie, Kevin Roden, Michael Veale, Stephanie Farquhar, Ryan Williams

We also greatly appreciate everyone from across the country who took the time to talk with us over phone and at conferences in Fall 2014 and beyond:

Seema Iyer (*Baltimore Neighborhood Indicators Alliance*), Kermit Wies (*Chicago Metropolitan Agency for Planning*), Brenna Berman (*City of Chicago*), Kevin Paris (*Community Advancement Network*), Vanessa Saria (*Community Advancement Network*), Chantal Stevens (*Community Indicators Consortium*), Erica Raleigh (*Data Driven Detroit*), Mark Abraham (*DataHaven*), Craig Helmstetter (*Minnesota Compass*), Ruth Hamberg (*Minnesota Compass*), Daniel O'Neill (*Smart Chicago Collaborative*), Kathy Pettit (*Urban Institute / National Neighborhood Indicators Partnership*), Steve Spiker (*Urban Strategies Council*)

TABLE OF CONTENTS

Executive Summary	1
Introduction	3
The Data Ecosystem Project	5
Vision for North Texas' Data Ecosystem	12
Essential Functions of the Data Ecosystem	13
Recommendations	21
Conclusions	24
End Notes	25

EXECUTIVE SUMMARY

North Texas faces a myriad of challenges as it moves through the second decade of the 21st Century. Affordable housing, rapid growth, aging infrastructure, poverty alleviation, children's health, and urban blight are just some of the many complex, interconnected problems that require an unprecedented amount of long-term action, cross-sector coordination, and development of a common agenda to address. Cities across the world have recognized that data is central to both understanding and acting on these complex issues and that the importance of making data accessible to a greater population is key to reaching these goals. A number of different strategies for achieving these goals have been adopted across the country, with each offering unique strengths and benefits.

The Data Ecosystem Project, sponsored by Communities Foundation of Texas, has worked with a diverse group of stakeholders to identify a system that can revolutionize the ways we collect, share, access, and use data in Dallas. Through research on national best practices, the Dallas Data Ecosystem Survey, and local stakeholder engagement the Data Ecosystem Project has identified an approach that organizations in Dallas and North Texas can work towards to improve the health and vitality of the regional data ecosystem.

The wealth of motivated organizations and individuals already working in Dallas and North Texas lends itself to collaborative solutions that benefit from the existing skills and expertise of the data ecosystem. The single greatest need is a centralized data catalog, library, or portal - where data from a variety of organizations, topics, and scales can be accessed by anyone and everyone. To do this, a mix of ancillary activities must also occur, from the creation of a governing body and governance structure to the development of topic specific cohort groups that encourage collaboration and participation between and across different parts of the ecosystem.

The Data Ecosystem Project has identified a data ecosystem model built around a number of primary and second functions. To implement these practices across the ecosystem, we recommend the following activities take place by Fall 2017:

1. Recruiting advocates from a variety of sectors and backgrounds to act as champions of the data ecosystem.
2. Acquiring preliminary funding to carry out a set of initial work items with oversight from a preliminary advisory team.
3. Forming an advisory team to help guide the development of the data ecosystem's final structure.

4. Conducting a data inventory to better understand the extent of publicly available data in North Texas.
5. Developing case studies that help demonstrate the value and potential of enhanced data accessibility.
6. Creating a business plan and securing multi-year funding for implementation.

Moving forward with these steps will support the successful implementation of a more robust data ecosystem. Whether the advisory team recommends and builds the case for a collaborative effort or the development of a university-driven data hub, taking the steps to make data more accessible will better enable the region to tackle the problems it faces. By maintaining the status quo, organizations must work harder and less efficiently to understand themselves, their communities, and make well-informed decisions.

INTRODUCTION

Making information resources accessible, discoverable, and useable by the public can help fuel entrepreneurship, innovation, scientific discovery, and community revitalization - all of which improve lives and create jobs.^{1 2} North Texas faces a wide range of issues that require cross-sector coordination, shared agendas, and reliable data to adequately address. While this may seem daunting, cities across the country are adopting innovative strategies that utilize a powerful combination of community partnerships and data driven decision making to tackle these and different issues.

Increasingly, these new approaches to problems solving rely on data. Technological advances, and innovative new applications of technology, have given rise to efforts that make data and government more transparent, accountable, and useful to residents. While data collection by cities and other government institutions is not new, these advances allow for unprecedented access to information by broader community audiences interested in making informed decisions and advocating for better service delivery, resource allocation, local policy, or a variety of other uses. Through the national Open Data movement, all levels of government are beginning to re-think how the wealth of information they collect is compiled and shared with constituents. As this expands to include more and more governments, it is important to evaluate the systems in place for accessing this data and information. This trend will begin to expand to other entities that collect and manage large amounts of information. Within the nonprofit community, philanthropic foundations have begun to demand more data from the organizations they fund to assess impact and evaluate programs.³

Improving access to data for those with the means to use it is only part of the equation - data relevant to the wider community should be available and easily digestible to as wide an audience as possible. Through this, communities can be empowered to advocate for their needs on the same page as their government and private businesses; small businesses can better access information vital to their success; and governments can better analyze their performance to improve services for constituents. Across the country, moves towards open data are transforming the way problems can be tackled by nonprofit, government, and businesses - from tools to better understand blight in Detroit, workshops to learn how to use open data resources, and even fellowships to promote data driven decision-making in New York City's Community Boards.^{4 5 6}

Addressing the major challenges of North Texas requires not only dedicated work within specific focus areas, but also a major focus on the underlying systems for collecting, accessing, and sharing data in the region. It is with data that we can make the transformative changes necessary to improve the lives of the regions 7.1 million inhabitants.⁷ Similar efforts across the country have better enabled local governments, nonprofits, individuals, and private businesses to work together to collect, share, and use data to better inform their decision-making processes.

To radically transform the systems of data collection, maintenance, analysis, and sharing that exist in North Texas, it is necessary to first understand major barriers to accessing data in the region, the extent of on-going activity, and which realistic, actionable steps can be taken. To do so, the Data Ecosystem Project (DEP) has involved a diverse group of stakeholders to identify the best approaches for improving access to data, enhancing existing data analysis and collection activities, and supporting continued collaboration to address the region's major challenges. In the following report, we provide a brief history of the DEP before documenting the major elements missing and needed in North Texas' data ecosystem and making final recommendations on how to achieve such an ambitious goal.

WHAT IS "OPEN DATA?"

"Open data is data that can be freely used, re-used and redistributed by anyone - subject only, at most, to the requirement to attribute and share alike." - Open Definition ⁸

"'[O]pen data' refers to publicly available data structured in a way that enables the data to be fully discoverable and usable by end users." - Office of Management & Budget ⁹

THE DATA ECOSYSTEM PROJECT

The Data Ecosystem Project (DEP), an initiative sponsored by Communities Foundation of Texas, seeks to revolutionize the ways we collect, share, access, and use data in Dallas. The Project, led by buildingcommunityWORKSHOP [bc], collaborated with a diverse group of stakeholders to understand how organizations in Dallas and North Texas are using data, the barriers they face when doing so, and ultimately how they can collaborate to build a vibrant, healthy local data ecosystem.

As we define it, the data ecosystem is the body of data collected and shared across the region (resources); the organizations and individuals responsible for collecting, distributing, maintaining, analyzing, and using data (actors); and the connections between those resources and actors. Similar to an ecological ecosystem, a data ecosystem is a complex web of relationships and dependencies between various actors and the resources they produce or consume. From the beginning of the project, it was understood that our local data ecosystem is unhealthy. While there are a wealth of actors, connections between actors and resources are sporadic and resources are of varying quality. Across the country, governments and private entities have worked to build robust ecosystems for the data community that in turn better enable the rise of data-driven decision-making at all scales.

Unfortunately, no two cities are alike in their activities or data needs. Ecosystems fit the need of their communities, however, through the DEP we have identified common themes seen in efforts to democratize access to data across the country. Each of these best practices represent a different way organizations were found to interact with data - such as data collection, analysis, or storytelling - and the way these activities relate to internal or external functions of the organization (Figure 1).

FIGURE 1: THE DATA CYCLE

Based on best practices identified through national best practice research, bcANALYTICS created The Data Cycle to better understand the relationships between raw data collection, analysis, storytelling, and community engagement..



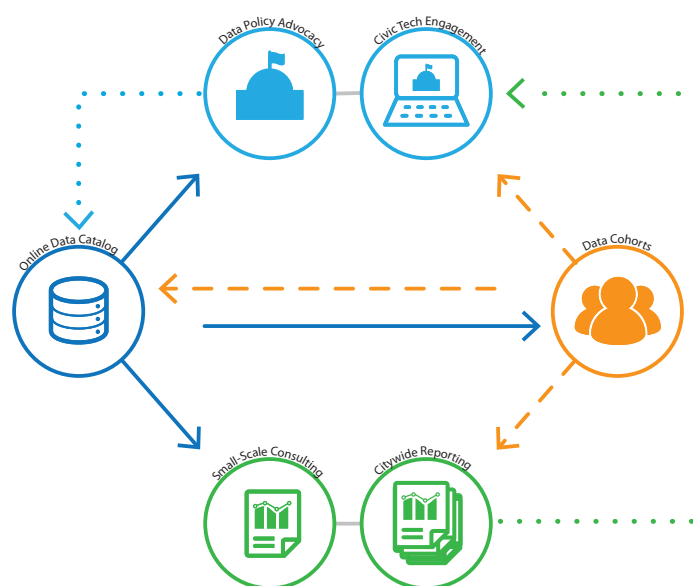
The focus of this initial best practice research focused on the nonprofit, philanthropic, and academic entities that serve as their city or region's data intermediary. These organizations often work with a variety of actors to enhance access to data and research needed across their communities. (For more information on these best practices, and some of the organizations used to understand these activities, see Appendix A.)

Best practices observed across the country helped better understand the types of projects and programs used to enhance and support data-driven decision-making nationally. To better understand the challenges and needs found in North Texas, the DEP hosted four key learning and engagement activities: a review of national best practices, a survey of the ecosystem, stakeholder-driven focus groups, and labs to identify workable solutions. Through each of these efforts, the DEP was able to gain greater understanding of the local data ecosystem and its needs.

A key finding of this work was the identification of 6 Key Functions present in robust data ecosystems across the country. Focusing on these 6 Functions, the DEP developed an initial model, seen in Figure 2, which was used as a discussion tool and further refined into a model to address many of the major issues found across the data ecosystem. In the following section, summaries of each phase of work build towards our final, working vision for the North Texas data ecosystem. The model that emerged through the DEP is presented in the following section.

FIGURE 2: 6 KEY FUNCTIONS OF DATA ECOSYSTEM

Activities that occur across an ecosystem were grouped into 6 functions. The functions work in relationship with each other to reinforce relationships between data, accessibility, capacity, and product.



National Best Practices

Between August 2014 and February 2015, the DEP researched the ways organizations and cities are leveraging data to bring positive changes to their communities. We spent time meeting, interviewing, and learning about the projects and products of leading organizations across the country. We developed simple case studies to understand the range of organizational structures and approaches in place across the country, the history of their development, and how the vast array of community-driven data initiatives can be incorporated into the functions of North Texas' data ecosystem.

Through this process we identified several best practices that emerged from the work of leading organizations to improve their communities, from building capacity for data use in organizations to providing access to machine-readable raw data (See Appendix A for more information). Learning from the various projects and initiatives is a vital part of the DEP. Each best practice provides a means for understanding the type of work necessary for transforming communities through data, and were used to develop and refine the Data Cycle, a tool for understanding how organizations engage with data and the internal and external connections of that activity (Figure 1).

The types of organizations we researched varied from collaborative entities with multiple partners to independent 501(c)(3) nonprofits and university research centers, with budgets from \$200,000 a year to almost \$2 million. The landscape of these data intermediaries and community indicator systems is also varied, with activity in cities and regions of all sizes. The National Neighborhood Indicators Partnership, part of the Urban Institute, has 40 affiliated organizations in 25 cities across the United States that all work to improve communities through data. No single approach appears to dominate this field of work, with the needs of the local community and stakeholders appearing to drive much of an organization's activity.

With this realization, the DEP set out to better understand the data ecosystem in North Texas with one overarching question: *Where is activity most needed to improve the health of North Texas' data ecosystem?*

Dallas Data Ecosystem Survey

In late 2014, [bc] launched the Dallas Data Ecosystem Survey to organizations across North Texas. Inspired by the work of the Smart Chicago Collaborative, the survey was designed to learn what data is used, how it is used, the software or technologies used to analyze it, and a wide variety of other information regarding the local data ecosystem.¹⁰ The survey was primarily distributed to

nonprofit organizations in the region, providing much needed information on the state of the nonprofit data ecosystem - from major barriers to accessing and using data to the types of data respondents want to access. Results from other sectors were limited, but the lack of resources inherent to nonprofit work helped identify barriers in organizations with the greatest need. The results of this survey helped identify areas where major focus is needed in improving our local data ecosystem.

Broadly, respondents have little time to dedicate to data work, and when they do they have considerable issues accessing the data they need. Beyond that, organizations may wish to incorporate data skills into their organizations, but have challenges hiring staff with the desired expertise in data collection, analysis, or management. To arm existing staff with the abilities needed to undertake this work requires better access to data and opportunities to gain new skills. This is particularly challenging for nonprofit organizations that cannot afford to buy data, hire consults, or conduct major research projects internally.

A variety of follow-up questions have been identified, and a second version of the survey would help expand our understanding of North Texas' data ecosystem. Understanding not only barriers, but areas of strength can help inform the direction of future ecosystem work.

Focus Groups

Stakeholders from Dallas' data community were convened for two focus groups in early 2015. These half-day sessions were used to confirm the current state of the data ecosystem, identify major barriers, and brainstorm ideas to improve it. Building off the earlier phases of work, the focus groups were intended to use the knowledge of local stakeholders to better understand the history of and possibilities for the data ecosystem.

In Focus Group 1, participants focused on three specific areas: identifying ways data can be used to inform a larger community; assessing the many gaps and challenges found in the local data ecosystem; and prioritizing where more timely solutions were needed to improve the data community. Participants engaged in various activities and developed a list of major challenges found across the data ecosystem.

In Focus Group 2, participants were asked to identify three types of solutions to address challenges in the data ecosystem: solutions entirely new to Dallas; solutions that exist in Dallas but may need some adjustment to become more effective; and successful solutions that exist in Dallas but would need to be scaled up or expanded to cover additional topics or geographies. The solutions identified helped us further

understand activity still needed in the ecosystem, and areas in the existing system where improvements are needed.

Labs

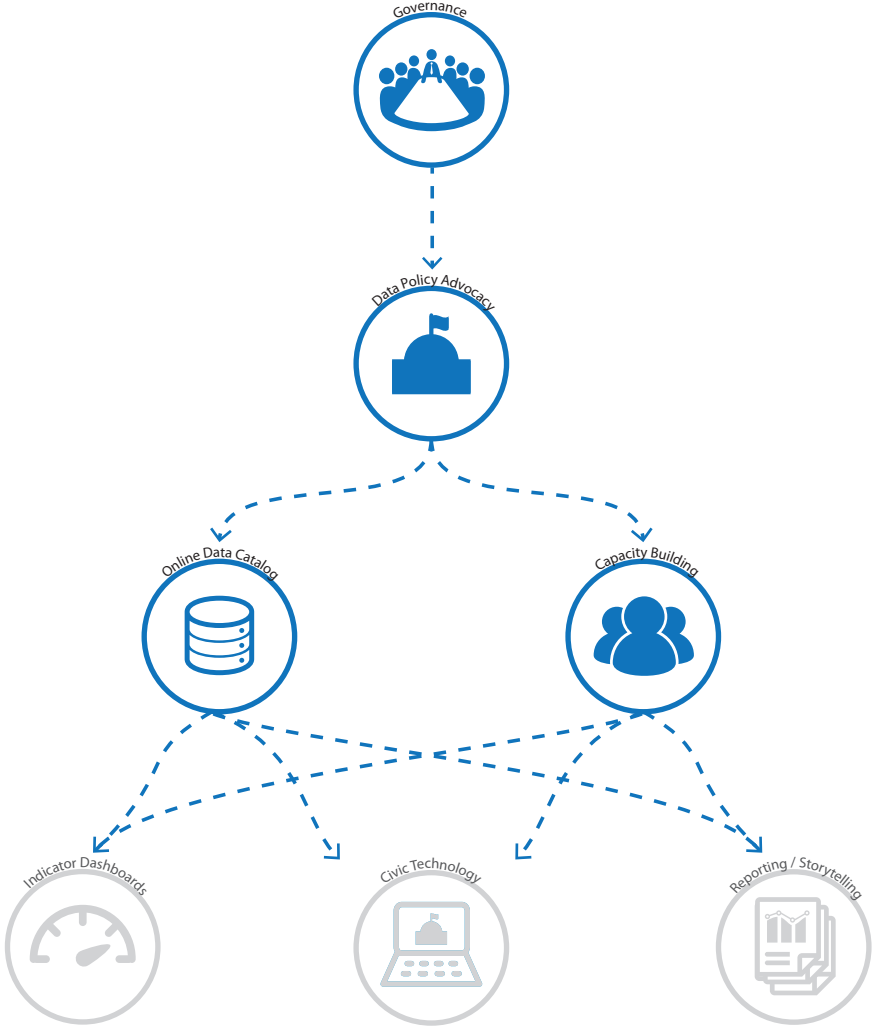
Through the earlier phases of work, we developed our initial model of the data ecosystem (the 6 Key Functions diagram on page 7). Using this model, we held two labs to review and develop a path forward. Each lab identified next steps and reorganized our understanding and use of the 6 Key Functions model, resulting in a more refined model discussed in the following section.

In Lab 1, participants were split into work groups based on the 6 Key Functions and asked to expand upon their definitions and to identify the extent of, challenges to, and key organizations related to the function. Through this activity, participants presented a need for governance within the ecosystem (a way for representatives from diverse sectors to establish priorities and encourage collaboration). Participants also began to refocus attention on the three functions that more directly enhance activity within the ecosystem - the Open Data Catalog, Data Policy Advocacy, and Data Cohorts. These functions are where changes are more beneficial and impactful for the ecosystem, whereas the remaining functions primarily relate to products derived from research and analysis.

This change within our 6 Key Functions model reflected a common sentiment that breaking silos of work and increasing access to data are paramount, but that there remains an additional level of organizing activity needed in the ecosystem. A second iteration of the model was developed from these concepts which introduced a dichotomy of activity within the ecosystem between primary and secondary functions (Figure 3).

An overwhelming theme from Lab 1 focused on the need for a cohesive strategy for prioritization across the ecosystem, both to jumpstart needed changes and to better coordinate and tackle issues that arise across the ecosystem in the future. With this in mind, establishing the approach to a governance mechanism became a major focus for Lab 2. To create the needed systems, participants identified a number of vital next steps that will establish a vision for the data ecosystem, build the case for changes for decision-makers, and determine necessary funding for the work.

FIGURE 3:DATA ECOSYSTEM VERSION 2



VISION FOR NORTH TEXAS' DATA ECOSYSTEM

From its inception, the DEP has operated under the assumption that data should be readily available and be used to drive positive changes in our communities. We believe that data plays a vital role in supporting those policy changes, that promoting equity in data access, and building capacity for data literacy and analytics skills is key to growing that role.

To that end, the DEP believes that the data ecosystem should work to build trust, collaboration, expertise, and accessibility for all possible users. Doing so requires participation from a wide variety of entities with different motivations, needs, and capacities. Our vision for the data ecosystem promotes these sentiments. The Data Ecosystem should:

1. Organize public data to be easily available and accessible to all audiences;
2. Create an infrastructure that drives collaboration and minimizes the need for direct competition;
3. Build a culture of excellence, trust, and inclusiveness through open sharing of best practices, data, and strategies for public good;
4. Consult place and issue specific expertise in the development of research, data initiatives, and use of sensitive information to be fully responsive to the needs of specific communities;
5. Improve the skills and capacities of organizations and individuals to both use and share data;
6. Support the local technology sector to accelerate and test civic innovations;
7. Engage the policy sector to be fully informed by data and research, and to be responsive to the needs of the data community.

ESSENTIAL FUNCTIONS OF A DATA ECOSYSTEM

Through our review of national best practices we observed similar activities occurring in cities across the country. The Data Ecosystem Survey helped grow our understanding of the barriers to data use in North Texas and stakeholder engagement activities further refined our understanding of activities needed to support a healthy data ecosystem in North Texas. Resources and actors within the data ecosystem are not lacking in the region - a wealth of data is collected and maintained by a wide variety of entities. The DEP has found, however, that the connections between resources and actors are weak. Strengthening existing connections and creating new ones will enable the growth of the actors and resources across our data ecosystem.

The DEP argues that a healthy and vibrant data ecosystem should be organized around a set of five primary functions that better enable the development of tools, products, and research within three secondary functions (Figure 4). These secondary functions occur regardless of any changes, but will not drive towards an improved data ecosystem on their own. Each primary function discussed in this report is meant to work across the data ecosystem, complement existing systems, and support greater access to data and information for all users.

This approach merges national best practices with the needs of the local data ecosystem (the *primary functions*) and products or actions that arise from the use of data (*secondary functions*), as identified through the ecosystem survey and key stakeholders. In the following section, each of the five primary functions are defined and our vision for each function is laid out. In contrast with the 6 Key Functions, the primary and secondary functions we discuss in this section represent types of activities needed to build a healthier data ecosystem in North Texas. The functions do not identify a means of executing needed work, but represent activities and relationships that must be built into the ecosystem.

PRIMARY FUNCTIONS

Each primary function facilitates a number of related activities that build and enhance a healthy data ecosystem. A governance element can prioritize strategies, establish policies for data standards, and work to emphasize the benefit of data and data analysis across North Texas; but other activities are needed to transform the way organizations are able to work when it comes to data. An open data catalog is needed to better access data, and open data policies are needed to ensure that data is shared through the catalog. At the same time those working in related fields can benefit from sharing their experiences with peers, and new opportunities for skill development grow capacity across the data ecosystem. Through a system that fosters and supports these activities, organizations and individuals can better conduct work that transforms the city (through secondary functions).

FIGURE 4: PRIMARY & SECONDARY FUNCTIONS OF THE DATA ECOSYSTEM



Governance - A governing body and/or governance structures that facilitates the open sharing of data and information between organizations and institutions, prioritizes activities, and promotes work occurring within the ecosystem.



Open Data Advocacy - A dedicated team that works to promote and encourage open data policies in local governments and fosters data sharing agreements between different entities (including the Open Data Catalog).



Open Data Catalog / Portal - A centralized and comprehensive web-based platform that documents publicly available datasets and resources from cities and agencies across North Texas.



Data Cohorts - Work groups organized around specific topics, shared work areas, or geographies that advise governance, seek to address barriers, share ongoing work, and discuss potential collaborations.



Capacity Building - Workshops, seminars, lectures, trainings, or courses that improve data literacy by offering resources and skill development in the areas of data collection, analysis, interpretation, or visualization.



Indicator Dashboards - Tracking performance on key socioeconomic, demographic, economic, or other measures to help communities understand themselves and their neighbors.



Reporting / Storytelling - The use of data to inform academic and professional research, as well as storytelling reinforced and informed by data.



Civic Technology - Promoting the use of publicly available data and information to build websites, applications, or tools to help citizens better connect to the information and resources they need.

Governance

The Governance function brings structure and consensus to the data ecosystem. Implementing and sustaining changes within the local data ecosystem requires a system of governance that can work across sectors, industries, and governments to increase data accessibility, awareness, and use in North Texas. Ecosystem governance will bring together representatives from a wide variety of organizations from different sectors, industries, and governmental or quasi-governmental entities. Major decisions needed to help shape the scope of the ecosystem will be determined by the governing body - from establishing data standards to setting priorities for collaboration across the data ecosystem. Revisiting these decisions and establishing new directions for the data ecosystem will be equally as important as technological changes enable new and different approaches to work across the data ecosystem.

As trust between data producers and consumers is built, expanded collaborative work will play a key role in guiding the development of standards for data collection, quality, accessibility and documentation. When stakeholders agree to these standards - across organizations and entities - consumers of data will be able to increase the scale and quality of their work and inform the systems used by data producers to collect and compile data. Achieving these goals would be much more challenging without some system of governance in place.

Open Data Advocacy

The Open Data Advocacy function brings attention and action to open data policies and their adoption across the region. Governmental and organizational policies often have a major impact on how data is collected and ultimately used, but in many cases there is no policy in place to shape how data can be accessed or shared by third parties. 38 city governments have adopted Open Data policies in the United States (there are more than 19,000 municipalities and 16,000 townships in the United States).^{11 12} Most of these policies were implemented in the wake of sweeping federal changes to data accessibility implemented by the Obama Administration starting in 2009, and as additional changes to federal open data policies occur state and local governments must begin to adopt similar practices.¹³ A dedicated team is needed to promote and support these types of policy changes across North Texas.

An Open Data Advocacy team would support the overall data ecosystem by negotiating data sharing agreements between entities, promoting open data policies, and developing relationships to help organizations access needed data. The team would focus a bulk of its work on developing data sharing

agreements or memoranda of understanding with data producers to either allow access through an open data catalog or for use in specific projects of other organizations. This can include sharing document templates that organizations can use to share their data and protect themselves from the misuse of data. Advocating for the expanded adoption of open data policies in governments across the region would be a secondary focus of the team. To do so, the team must demonstrate the value of open data locally, in coordination with the Data Cohorts and Governance functions.

To succeed at both of these roles, the team must build strong relationships with data producers to help better advocate for the needs of consumers - be they nonprofits, universities, or small businesses. The Open Data Advocacy team would be the go-to group for finding and accessing needed data that is not publicly available.

Open Data Catalog / Portal

Finding locally relevant datasets is a major challenge in North Texas, and a means to address this is one of the region's greatest needs. Creating a system where publicly available data can be easily found, and data needs can be identified, is a cornerstone to a healthy data ecosystem. In execution, this should be a single, centralized location that will minimize users' frustrations when looking for resources that are often difficult to identify or access. A core team or entity is needed to manage this catalog and ensure that resources are up to date, accurate, and complete. This design would not include access to personally identifiable data, and all necessary steps must be taken to ensure that data cannot be used to identify or endanger any person.

Ultimately, the Open Data Catalog promotes greater transparency by arming researchers, business leaders, government officials, students, and community members with access to data and information regarding their businesses, communities, and lives. As changes continue to occur within the data ecosystem and more data becomes publicly accessible, a centralized data catalog must be able to adapt to the needs of the ecosystem. Long-term needs of the data catalog should be consistently evaluated to keep the region on pace with technological change.

These changes, however, are not simple. City employees are already overburdened, and adding new job responsibilities require funding and additional staff. Identifying a team or organization to maintain and build this catalog is paramount. The team can take on a variety of tasks, determined by the governing body, that will expand access to data across the region and better

enable the creation of secondary products, such as community indicator platforms and mobile applications.

Data Cohorts

Data Cohorts bring awareness and collaboration to defined areas of the data ecosystem, and provide expertise and on the ground knowledge to the governance function. The Data Cohorts will primarily be formalized through the creation of cohorts for different topic areas, geographies, or sectors that are able to build and support strong relationships between practitioners. Providing opportunities for data analysts, researchers, government officials, and community members to engage with one another, share work with, and identify common challenges within a specific domain (such as education, health, the environment) will strengthen the ability of the individuals who make up the data ecosystem.

The ecosystem's priorities are established through governance, but input from those working closely with data will help identify where changes are needed. The Data Cohort function will act as a means to monitor activity and health within the data ecosystem. Organizations or sectors that struggle to access or use data will find opportunities to share their issues with relevant cohorts, identify possible solutions, or drive action across the ecosystem by informing the governing body. If problems are consistently identified within a cohort or across several cohorts the Governance function might work to address the solution or bring attention to the issue from a higher level.

Capacity Building

Capacity Building brings expanded learning opportunities to those interested in engaging with the data ecosystem. Building a healthier data ecosystem in North Texas requires more than increasing access to data. At one end of the spectrum, increasing the understanding of data's importance and uses can expand data literacy across the region, and at the other, current data analysts, data scientists, researchers, and technologists can expand their skills or keep up with changing technologies. This involves increasing community awareness of existing resources, how they can be used, and why they matter, but also a better understanding of how to critically evaluate data and analytic products. Consistent opportunities must be available to those with different skillsets and backgrounds, ranging from workshops on using data software to small courses covering best practices for data collection.

While Capacity Building opportunities are needed for every level of user, it is likely that initial activities will focus on nonprofits and small businesses.

Developing a wide range of data and analytics techniques can be incredibly important to those sectors, especially as foundations and investors demand to see an ever-greater amount of data and information to justify investments. Avenues for gaining these skillsets exist across the region, but identifying opportunities for new programs or approaches will be just as important as ensuring the continuation of existing programs.¹⁴ Stakeholders from organizations currently providing these types of opportunities should be identified to establish gaps in the region's current offerings, and determine how best to expand offerings to new audiences and topic areas. This might emerge through new providers, guides created by Data Cohorts, or incorporating needed activities into existing programming.

SECONDARY FUNCTIONS

For the most part, secondary functions emerge as the result of data and are found in places with both vibrant and limited ecosystems. With increased access to data, it is simpler to conduct research, build applications, make important organizational decisions, and understand one's community. In most scenarios, however, data awareness is increased through public demonstrations of its use. As such, much of the DEP has focused on ways to embed this support into the ecosystem to help secondary functions thrive across the region.

Indicator Dashboards

Creating and tracking appropriate metrics to measure community health and performance is a key role of certain actors within a data ecosystem. Taking data from a variety of sources and creating platforms to compare and explore data across cities, neighborhoods, or governing districts is a great asset for any community. With enhanced access to data, it becomes possible to generate timely insights on a neighborhood to drive decision-making at all scales. These systems currently exist, but can continue to improve only through better data sharing practices and communication across the ecosystem.

Reporting / Storytelling

Academics, consultants, journalists, and other researchers are adept and agile when it comes to finding the right data for their needs. Unfortunately, this often comes with a large price tag related to data purchase requirements or staff time needed to compile and clean data. This high cost of data and research provides a challenge for small organizations and entities that wish to undertake these tasks on their own. Better access to data enables a wider range of data analysis and research activities across the ecosystem, and contributes to a better community-wide understanding of the issues affecting communities across the region.

Civic Technology

The rise of open data has reshaped how technology can enhance civic and public participation. These innovative new applications are able to perform a wide variety of services, from streamlining the process of paying a speeding ticket to connecting citizens to quality information about their community and health needs.^{15 16} Without quality access to machine-readable or API-enabled open data, civic innovators may choose to seek regions where access to this type of data is possible. Enabling resources for entrepreneurial actors can help stimulate the creation of small-businesses across the region.

RECOMMENDATIONS

The DEP believes that the North Texas data community should work together to create a healthy, vibrant data ecosystem with improved systems of collecting, accessing, and sharing data. Achieving this goal will take a concerted effort and investment from a number of individuals and organizations who recognize the value of data and are committed to seeing these changes occur. The DEP strongly believes that vested stakeholders should work to create a collaborative entity that will develop and oversee work falling under each of the 5 Primary Functions discussed in the previous section.

To do so, we recommend a series of short term actions that will carry this work forward and lead to a more vibrant data ecosystem in North Texas:

1 - Acquire preliminary funding to carry out a set of initial work items

Everything has a price. To make progress towards a healthier data ecosystem, funding is needed in the short-term to carry out a specific series of tasks outlined below. Identifying and acquiring funding to see this work through is vital to its success, as each of the remaining recommendations begins to formalize different aspects of the primary functions (Governance, Open Data Advocacy, and the Open Data Catalog, in particular).

2 - Recruit advocates from a variety of sectors and backgrounds to act as champions of the data ecosystem

Building the case for Open Data and promoting the great work conducted in the local data ecosystem requires the attention of individuals and organizations of all types. Identifying existing champions of open data, technology, and transparency is key. These individuals should work to build trust, goodwill, and awareness of the data ecosystem and advocate for needed changes.

3 - Form an advisory team to help guide the development of the data ecosystem's final structure

Key stakeholders, both organizations and individuals, should be approached to participate in a preliminary advisory team. This group will facilitate the implementation of governance and funding structures for the data ecosystem and act on the short-term work items that must occur prior to official funding (outlined in these recommendations).

To encourage collaboration across the data ecosystem, members of the advisory team must come from a diverse group of organizations and entities. A balance between data producers and data consumers must be found to incorporate knowledge from different perspectives within the ecosystem.

4 - Conduct a data inventory to better understand the extent of publicly available data in North Texas and create a preliminary Open Data Catalog

A wealth of data is available for different areas across North Texas, but understanding the full scope of data collected across the region is a daunting task. Building a working inventory of locally relevant data is a key task of the preliminary advisory panel, and will help shape and form the content available in the Open Data Catalog. The inventory process should yield a preliminary Open Data Catalog that is publicly accessible.

5 - Develop case studies that help demonstrate the value and potential of enhanced data accessibility

There are many areas where the lack of data, or lack of access to data, can limit the opportunities of small businesses, governments, nonprofits, and researchers. The advisory panel and other advocates should work to demonstrate the challenges of using data in key topic areas and bring attention to the problem as a whole. Identifying and documenting pertinent examples of these barriers and what is possible when they are removed will help clarify the need for changes to the status quo.

6 - Create a business plan and secure multi-year funding for implementation.

A business plan and multi-year funding will establish this effort in Dallas. The creation of a new organization or collaboration will provide the needed resources and dedicated focused needed to carry out the core body of work. This process will establish the needed resources (staffing, monetary, technological) and identify the full scope of this entity's roles, responsibilities, and work items. This is a critical step in moving this initiative forward and key to the creation of a new entity in the North Texas data ecosystem.

CONCLUSION

The Data Ecosystem Project has sought to understand and improve the health of North Texas' data ecosystem since 2014. We have seen examples of amazing work in data ecosystems across the country and are determined to see similar changes occur in our own communities. Organizations across North Texas are hungry for data. To address this appetite, governments, nonprofits, businesses, universities, and community groups must work together to bring about the needed changes.

Recommendations laid out in the previous section offer an approach to implementing the 5 Primary Functions through the creation of a singular data intermediary or a collaboration with actors taking on various roles based on expertise and need. The ultimate form of this intermediary must be determined by the broader community, but must build trust across the ecosystem to strengthen its activities and the changes it must oversee. Without these changes, it is difficult to imagine any change within North Texas' data ecosystem. Sporadic, intermittent changes will occur and slowly begin to spread to new areas, but this will only reinforce the challenges faced by actors in the status quo.

We have developed a model that supports existing work, creates vibrant networks of practitioners, enhances access to data for all audiences, and encourages collaboration across sectors and between actors. The 5 Primary Functions outlined in this report offer a way to tackle barriers that stem from similar, if not related, issues. By listening to stakeholders and organizations across the region we know that changes are needed. Where the Open Data Catalog and Open Data Advocacy functions improve the ways that individuals and organizations find and access data, the Data Cohorts and Capacity Building functions empower individuals by expanding their professional networks, increasing awareness of other work, and providing opportunities for gaining new skills. To oversee activities in each area, the Governance function incorporates knowledge from across the ecosystem to set priorities and drive forward some of the most needed changes. In concert, these activities seek to overcome the issues of the status quo and democratize access to data across North Texas.

END NOTES

1. Chui, M., D. Farrell, and S. Van Kuiken. 2013. Generating Economic Value through Open Data. B. Goldstein, L. Dyson, and Abhi Nemani (Eds.). *Beyond Transparency: Open Data and the Future of Civic Innovation* (pp 163 - 172). San Francisco, CA: Code for America Press.
2. Gordon, E., J. Baldwin-Philippi. 2013. Making a Habit Out of Engagement: How the Culture of Open Data Is Reframing Civic Life. B. Goldstein, L. Dyson, and Abhi Nemani (Eds.). *Beyond Transparency: Open Data and the Future of Civic Innovation* (pp 139 - 150). San Francisco, CA: Code for America Press.
3. Vasquez, S. and P. Barry. 2014. Everything Old Is New Again: Building Nonprofit Capacity in the Age of Big Data. Federal Reserve Bank of San Francisco, Urban Institute, N. Cytron, K. L. S. Pettit, and G. T. Kingsley (Eds.). *What Counts: Harnessing Data For America's Communities* (pp 319-334). Washington, D.C.: Urban Institute.
4. Motor City Mapping, a project from Data Driven Detroit and Loveland Technologies, created a publicly available, parcel-by-parcel assessment of blight across Detroit. For more information: <http://motorcitymapping.org>
5. The Chicago Data Learnathon was a two-day event to help data analysts and data scientists build their skills and learn how to better access and use the wealth of data available in Chicago and Cook County. For more information: https://sway.com/JQ82_ZTTh3ZUWyeU
6. The NYC Civic Innovation Fellows program aims to build data literacy in community organizations across New York City. To do so, twelve fellows are assigned to a Community Board in Manhattan to build open data practices and knowledge. For more information: <https://beta.nyc/programs/nyc-civic-innovation-fellows/>
7. U.S. Census Bureau Population Estimate for the Dallas-Fort Worth-Arlington Metropolitan Statistical Area in 2015.
8. The Open Definition 2.1. (n.d.) Retrieved April 20, 2016, from <http://opendefinition.org/od/2.1/en/>.
9. OMB Memorandum M-13-13. Open Data Policy - Managing Information as an Asset. <https://www.whitehouse.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf>
10. For more information on the Smart Chicago Collaborative's work on Chicago's data ecosystem, and to learn more about SCC, visit: <http://www.smartchicagocollaborative.org/work/ecosystem/chicago-school-of-data/>
11. Sunlight Foundation. (n.d.). Retrieved April 20, 2016, from <https://sunlightfoundation.com/policy/opendatamap/>

12. As of the 2012 U.S. Census of Governments, there were an estimated 19,519 municipal governments and 16,360 township governments in the United States. Data compiled by the U.S. Census Bureau: <http://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?src=bkmk>
13. These broad changes continue to occur at the Federal level, as seen with the continued focus on open data initiatives by The White House and the introduction of the bi-partisan Open, Public, Electronic, and Necessary (OPEN) Government Data Act in April 2016.
14. For just two local examples, Per Scholas (<http://perscholas.org>) and The Iron Yard (<http://theironyard.com>) both offer classes in Dallas to different populations that wish to build digital and technology literacy.
15. For more information on the CourtBot and ATLCourt applications, more information can be found at Code for America (<https://www.codeforamerica.org/why-government/our-partners/atlanta-ga>).
16. Smart Chicago Collaborative, the Chicago Department of Public Health, and a number of area hospitals have worked to compile a variety of information about Chicago neighborhoods health-related issues. For more information visit <http://www.chicagohealthatlas.org/about>.

