

# **Better the Data You Know: Developing Youth Data Literacy in Schools and Informal Learning Environments**

Erica Deahl  
B.A. Art, Yale University, 2007

Submitted to the Program in Comparative Media Studies/Writing in partial fulfillment of the requirements for the degree of Master of Science in Comparative Media Studies at the Massachusetts Institute of Technology.

June 2014

©Erica Deahl, 2014. All rights reserved. The author hereby grants to MIT permission to reproduce and to distribute publicly paper and electronic copies of this thesis document in whole or in part in any medium now known or hereafter created.

**AUTHOR** .....

Erica Deahl  
Department of Comparative Media Studies  
May 9, 2014

**CERTIFIED AND ACCEPTED BY** .....

Heather Hendershot  
Professor of Comparative Media  
Director of Graduate Studies, CMS



# Better the Data You Know: Developing Youth Data Literacy in Schools and Informal Learning Environments

Erica Deahl

Submitted to the Program in Comparative Media Studies/Writing on May 9, 2014 in partial fulfillment of the requirements for the degree of Master of Science in Comparative Media Studies.

## ABSTRACT

We live in an era of unprecedented growth in the use and impact of data. While large institutions are using data about individuals to drive decision-making, small organizations and civic hackers are using open and public data to innovate for social good. Meanwhile, the educational and informational resources to enable individuals to understand this data remain scarce. Individuals and communities are often unaware of the data being collected about them, the data they are contributing, and the ways in which that data is being used. Although the Open Data movement has given some individuals a new opportunity to interface with data directly, the public at large lacks the skills and knowledge to take advantage of this opportunity. This thesis argues that we need to support the public — especially youth — in developing data literacy, so that they are equipped to think critically and ethically about data. I make this case in four ways. First, I contextualize the need for data literacy by describing the historical evolution of institutional data collection practices, contemporary uses of data that have had a profound impact on institutions and individuals, and the potentially problematic consequences of data modeling. Second, I propose a definition of “data literacy,” situate the concept within the landscape of new media literacies, and describe settings, methodologies, and tools that can be used to support it. Third, I analyze two data literacy initiatives that enable youth to use data to investigate and address real-world issues: one in an informal learning environment, *Young Rewired State’s Festival of Code*, and the other in a public school, *City Digits: Local Lotto*. Fourth, I analyze the challenges facing data literacy initiatives — from the constraints of the public school environment, to the challenges of reaching diverse audiences and supporting open-ended learning. I propose three design principles to guide researchers, educators, and practitioners in shaping future data literacy initiatives.

## THESIS SUPERVISOR

Heather Hendershot

Professor of Comparative Media; Director of Graduate Studies, CMS



# Acknowledgments

I would like to start by thanking my advisor, Heather Hendershot. Heather, thank you for supporting me throughout this process and for challenging me to write a better thesis.

I would also like to sincerely thank Sarah Williams, my thesis reader and mentor at the Civic Data Design Lab. Sarah, thank you for giving me the opportunity to work with you and for your advice and support along the journey. I couldn't have gotten here without you.

Thank you to my other collaborators from the City Digits project: Laurie Rubel, Vivian Lim, Valeria Mogilevich, Pema Domingo-Barker, Chris Rhie, Lauren Shookhoff, Mathew Sullivan, Jose Ojeda, Vikash Dat, and the CDDL team. And thank you to the many wonderful high school students we worked with, who were the inspiration for this thesis.

Many thanks to Emma Mulqueeny, Kaitlin Dunning, Daniel O'Neil, Jeff McCarter, Andrea Hart, Joshua Prudowsky, Milena Marin, and Francois Grey for the opportunity to learn about your exciting work in the field, and to Philipp Schmidt, Mitch Resnick, Natalie Rusk, and Scot Osterweil for your ideas along the way.

I am extremely grateful that I was able to spend the last two years as part of the Comparative Media Studies community. I met some of my closest collaborators and friends in the CMS class of 2014: Denise Cheng, Julie Fischer, Alexandre Goncalves, Jason Lipshin, Lingyuxiu Zhong — and special thanks and hugs to Eduardo Marisca for your input and encouragement. Thank you also to William Urrichio, Jim Paradis, Fox Harrell, and the other CMS faculty for creating an inspiring environment to work in, and to Shannon Larkin for your endless supply of chocolates and Kleenex.

I'd also like to thank my family, especially my mom, Lora Deahl, for reading drafts, giving me pep talks, and providing much needed support and understanding. Thank you also to Aditi Mehta and Ankita Deshpande for putting up with me, and to Zoey the cat for sitting on my lap and purring through many hours of thesis writing.

Finally, I would like to thank Rodrigo Davies. Rodrigo, thank you for having faith in me, for dropping everything to help me, for reminding me what I was working for when things were hard, and for keeping me smiling.



# Contents

<b>INTRODUCTION .....</b>	<b>9</b>
<b>CHAPTER 1: The Growing Institutional and Individual Impact of Data .....</b>	<b>15</b>
1.1 The Growth of Data Collection .....	16
1.1.1 The Census: Statistical Data Collection in the United States .....	16
1.1.2 From Planned Data Collection to Ubiquitous Data Collection .....	19
1.2 The Impact of Top-Down Data Management .....	20
1.2.1 Data Modeling from Systems Theory to the Smart City .....	21
1.2.2 Problems of Data Modeling .....	24
1.3 A Counter Move: Open Data and Bottom-Up Data Innovation .....	27
1.3.1 Institutions Move Toward Open Data .....	28
1.3.2 Grassroots Data Collection and Visualization .....	29
1.3.3 Limitations of the Open Data and Grassroots Data Movements .....	31
1.4 The Public Needs Better Data Literacy .....	34
<b>CHAPTER 2: Data Literacy: Definition and Methods .....</b>	<b>39</b>
2.1 What is Data Literacy? .....	40
2.1.1 Data Literacy Competencies .....	40
2.1.2 Situating Data Literacy .....	43
2.1.3 Benefits of Data Literacy .....	45
2.2 How Can We Support Data Literacy? .....	48
2.2.1 Educational Technologies .....	50
2.2.2 Out-of-School Learning Models .....	52
2.2.3 Data Literacy in Schools .....	54

<b>CHAPTER 3: Two Initiatives for Building Socially Embedded Youth Data Literacy .....</b>	<b>61</b>
3.1 Young Rewired State's Festival of Code .....	64
3.1.1 Background .....	65
3.1.2 Program Overview .....	66
3.1.3 Outcomes .....	71
3.1.4 Conclusion .....	76
3.2 City Digits: Local Lotto .....	77
3.2.1 Background .....	78
3.2.2 Program Overview .....	81
3.2.3 Outcomes .....	89
3.2.4 Conclusion .....	92
3.3 Conclusion .....	93
<b>CHAPTER 4: Data Literacy Challenges and Design Principles .....</b>	<b>97</b>
4.1 Institutional Constraints of Public Schools .....	99
4.2 Diversity and Access Challenges .....	104
4.3 Pedagogical Challenges .....	108
4.4 Conclusion .....	92
<b>BIBLIOGRAPHY .....</b>	<b>115</b>

# Introduction

I'm Zak, and I'm 8.... Some scientists were tracking students in the USA to like find their interactions. We used that open data and got like these lists and programs to simulate disease spread, and [showed] it by turning little dots red to show that they're zombies. Basically, on the computer, it's like there's kind of nothing else in the world like it, basically. It's kind of like a new place, where you can make whatever you want.<sup>1</sup>

Zak first attended Young Rewired State's annual data hackathon in 2012 when he was seven years old. He did not have much experience coding software. At the event, he was inspired by the caliber of projects other young people were creating with data, and he was determined to return the following year to win one of the top prizes. He spent the next year experimenting with data and learning a variety of programming languages. When he returned in 2013, he recruited a team of talented seventeen and eighteen year olds with whom he collaborated to develop Contag.io, an app modeling the spread of a virus within social groups in a high school. The app won the prize for the Best Example of Code.

Zak is an example of someone who is *data literate*. He knows how to find, analyze, and create practical applications using data. His experience working with data has given him the confidence and motivation to seek out further interactions with data, undeterred by the complexity of the tasks involved. His story is proof that, given the right opportunities and resources, young people can become excited, engaged, and fearless in their interactions with data—and furthermore, that they are capable of complex and innovative work.

Yet Zak's experience is unusual. Few children his age—indeed, few adults—have interacted with data, let alone feel empowered in their interactions. They are typically the subjects of data, not the experts: governments collect statistical and demographic information about them while corporations amass data about their consumption preferences. This is especially true of young people, who have even fewer opportunities than adults to learn about data, and whose own student data is carefully tracked to assess learning. Most people regard data with the same passivity that they once regarded computers—as a technical artifact that is the property of experts and whose impact is difficult to challenge. “Data literacy” is not a term that has been widely used—and when it is, it most often refers to college and professional level education.

I am producing this thesis at a moment when this disconnect is more visible than ever. On the one hand, critics have pointed out the ethical challenges inherent in data-driven decision-making. In the news, we hear disturbing stories about how much corporations know about us, such as Target finding out about a teenage girl’s pregnancy before her father did based on her shopping history. On the other hand, data modeling has become a prominent practice—and data scientists are in high demand—because it has proven to be so profitable.

At this moment of tension, I believe we have an opportunity to shift the dialogue about data through youth education. Teaching young people to work with data will both enable them to understand the ethical dilemmas of data, and therefore to correct for some of the errors of the past in their future work, while also enabling them to participate in and advance a new field. In addition, we have the opportunity to create spaces for applying data literacy skills to social issues, rather than focusing on commercial purposes. Finally, we can change the perception that data exploration is too complex for young people. One inspirational model is the recent evolution of youth computer science learning. We have seen a large population of young people become expert programmers thanks to resources

such as MIT's Scratch, which lowers the barriers to participation and enables young people to create personally meaningful projects. We are in a moment when young people are beginning to be able to do this with data.

I was inspired to undertake my thesis on this topic after working on City Digits, a project that introduces high school students to data within the context of a social justice themed math class. In the course of this project, I learned how challenging data literacy initiatives can be to deliver, but I also witnessed the potential data literacy initiatives have for engaging youth and getting them interested and involved in social issues. This experience led me to understand the importance of defining goals and methods clearly. It also convinced me of the importance of this endeavor, because I realized that data literacy has broader implications for youth civic engagement. In my research, I wanted to explore the multiple ways in which researchers, educators, and practitioners are approaching data literacy—how initiatives are supporting data literacy, what challenges they face, and where we should look for inspiration and direction as we move forward.

My work is guided by four central research questions:

1. How are initiatives in public schools and informal learning environments supporting youth in developing data literacy?
2. What attributes characterize successful initiatives?
3. What are the challenges and limitations of these initiatives?
4. What principles should guide the design and implementation of future data literacy initiatives and the digital technologies that support them?

I explore these questions in four chapters.

In Chapter 1, "The Growing Institutional and Individual Impact of Data," I describe the new uses of data that have had a profound impact on institutions and individuals and the problems with data modeling that have made it important for the public to become data literate. The chapter is organized into four sections. The first section, "The Growth of

Data Collection,” describes the historical evolution of data collection practices, focusing on statistical data collection in the United States, the shift in recent years from planned data collection to ubiquitous data collection, and the evolution of data collection and generation as a profitable business model. I argue that the public needs better data literacy in order to understand how their personal data is being used. The second section, “The Impact of Top-Down Data Management,” describes how governments have relied on predictive data modeling and explores the ethical dilemmas and pitfalls inherent in data-driven decision-making. I argue that the public needs better data literacy in order to understand and question these decisions. The third section, “A Counter Move: Open Data and Bottom-Up Data Innovation,” describes how the Open Data movement has given individuals a new opportunity to interface with data directly. Here, I explain that the public lacks the skills and knowledge to take advantage of this opportunity. In the final section, “The Public Needs Better Data Literacy,” I argue that we need to support the public—especially youth—in developing data literacy, so that they are equipped to think critically and ethically about data.

In Chapter 2, “Data Literacy: Definitions and Methods,” I break down the definition, context, and objectives of data literacy and describe the range of methods that can be used to support youth in the development of data literacy. The chapter is divided into two sections. The first section, “What is Data Literacy?” describes the cultural and social implications embedded in the term “literacy,” proposes a working definition of data literacy and its range of objectives, and situates the concept of data literacy within the landscape of new media literacies. The second section, “How Can We Support Data Literacy?” explores the settings, methodologies, and tools that can be used to support data literacy. I describe my theories of learning and aspirations for data literacy initiatives, exploring the possibilities for project based, student centered, and culturally relevant learning. I end by mapping out the range of technologies, out-of-school programs, and in-school programs that currently operate in this field.

In Chapter 3, “Two Initiatives for Building Socially Embedded Youth Data Literacy,” I focus on two data literacy initiatives that enable youth to use data to investigate and address real-world issues, one in an informal learning environment and the other in a public school. The first case study describes *Young Rewired State’s Festival of Code*, a weeklong hackathon where youth create websites, apps, and algorithms using open government data. The second case study describes *City Digits: Local Lotto*, a high school math curriculum in which students collect and analyze data to investigate local social justice issues. The case studies describe each project’s background, curriculum design, implementation, and the challenges and successes experienced by the participating youth and educators.

Finally in Chapter 4, “Data Literacy Challenges and Design Principles,” I analyze the challenges facing data literacy initiatives and define priorities for the design of future data literacy initiatives. The chapter is organized into three sections, each of which describes a challenge and proposes a corresponding design principle. The first section, “Institutional Constraints of Public Schools,” describes the challenges of aligning a data literacy initiative with public school curriculum requirements, getting the support of teachers and administration, and enabling student-centered learning within the constraints of the school day. The second section, “Diversity and Access Challenges,” describes the challenges both in and out of schools of engaging diverse audiences from demographics traditionally underrepresented in the fields of programming and computer science, and describes the resource challenges which make replicating data literacy initiatives difficult for schools and communities with limited resources. The third section, “Pedagogical Challenges,” describes the barriers educators face as they adopt new content, methods, and technologies. It also describes the pedagogical challenges of supporting student agency, tailoring learning to diverse audiences, and replicating initiatives that are personalized for individuals or customized for specific communities. To address these challenges I propose three design

principles to guide researchers, educators, and practitioners in shaping future data literacy initiatives: 1. Create a data literacy ecosystem in and out of schools; 2. Design for a range of levels, communities, and learning styles; 3. Create training resources for teachers and support open-ended learning.

## NOTES

- 1 *Zak, 8 – Festival of Code Interview #yrs2013*, 2013, [http://www.youtube.com/watch?v=LKy2O0Ef5lI&feature=youtube\\_gdata\\_player](http://www.youtube.com/watch?v=LKy2O0Ef5lI&feature=youtube_gdata_player).

## CHAPTER 1

# The Growing Institutional and Individual Impact of Data

We live in an era of unprecedented growth in the use and impact of data. Large institutions are using data about individuals' spending habits and content preferences to drive decision-making, and small organizations and civic hackers are using open government records and public biographical data to innovate for social good. Meanwhile, the educational and informational resources to enable individuals to understand this data remain scarce. Individuals and communities are often unaware of the data being collected about them, the data they are contributing, and the ways in which that data is being used. In this chapter, I argue that the public's understanding of data has failed to develop in proportion to the increasing scope and impact of data, leaving citizens ill equipped to understand and participate in processes that affect them.

This chapter is organized into three sections. In the first section, "The Growth of Data Collection," I describe how data collection has become increasingly automatic and ubiquitous—stitutions are collecting an unprecedented amount of data, while individuals are the passive subjects of data collection, often unaware of what data is being transmitted from their computers and mobile devices. I argue that the public needs data literacy in order to understand what is happening with their data. In the second section, "The Impact

of “Top-Down Data Management,” I describe how governments and corporations are using data to drive decision-making, a practice that often improves efficiency but is prone to human error. I argue that the public needs data literacy in order to understand and question these decisions. In the third section, “A Counter Move: Open Data and Bottom-Up Innovation” I describe how the open data movement and bottom-up data collection practices have provided individuals with new opportunities to actively collect and use data. I argue, however, that most individuals lack the skills and knowledge to participate in these movements; therefore, the public needs better data literacy.

## 1.1 The Growth of Data Collection

Governments have long collected statistical data about individuals in order to provide services, collect taxes, and enforce laws for a growing and changing population. A famous early example of this practice is the Domesday book, a manuscript surveying large parts of England and Wales that was completed in 1086 in order to ascertain the fiscal rights of King William I of England. The survey, which records 13,418 places, was conducted by a group of royal officers and was copied by hand on parchment.<sup>1</sup> From Domesday to the present, we have witnessed a gradual increase in the amount of data that is collected and the speed at which data is collected, processed, and analyzed.

### 1.1.1 The Census: Statistical Data Collection in the United States

The United States census is one longstanding data collection practice that illustrates that this shift in the scope and speed of data collection has been unfolding for hundreds of years. The census also shows that despite changing methods, there is a resemblance between the challenges and ethical concerns that characterized data collection in the eighteenth century and those we face today. In particular, it establishes a precedent for individuals’ lack of

agency in the collection and use of their own personal data.

First, the scope of information that the government collects about individuals has dramatically increased. The first US census was conducted in 1790 and asked only for the name, age, gender, and race of each member of the household, including slaves, in order to facilitate taxation and military recruiting.<sup>2</sup> The fact that a provision for a census was written into the original Constitution reveals that leaders understood the value of data collection; yet the limited scope of questions suggests that they did not yet have the methods, objectives, or needs to collect and analyze extensive results. Over the years, the scope of data collected increased: the 1810 census began collecting economic data on manufactured goods; the 1850 census collected social statistics on taxes, education, crime, and estate value; and the 1940 census contained additional questions on internal migration, veteran status, and number of children born alive, information that helped measure the impact of the Great Depression.<sup>3</sup> This gradual increase in the amount of data collected reflects an emerging understanding of how to use data to assess the nation's economic and social conditions—and a growing technological capacity to do so.<sup>4</sup>

Second, new technological innovations, developed in response to the challenges of measuring a growing population, have increasingly automated the processes used to collect and analyze data. The greatest challenge for early census workers was data tabulation—the processing of census data collected in 1880, for instance, was not completed until 1887, just three years before the next census.<sup>5</sup> To address this issue, Herman Hollerith, a Census Bureau employee, developed an electronic tabulating machine that was first used during the 1890 census. To process a card, "the operator simply placed it on a rubber pad, beneath which lay dozens of tiny cups of mercury, and with a handle lowered a swinging array of metal pins. On a panel facing the operator, four rows of ten clocklike dials represented the various data items encoded on the card."<sup>6</sup> Hollerith's electronic tabulator had an enormous impact on the efficiency of census tabulation. More recent innovations have also increased

the efficiency of data collection as well as data tabulation. In 1960, the Census Bureau began mailing a census form rather than sending enumerators door-to-door; and in the 1990s, the Bureau began using computer-assisted interviewing and electronic reporting. These innovations and others have helped increase the fidelity and speed of data collection and analysis. They also shape a larger narrative of data innovation: in 1896, Hollerith left the Census Bureau to found the Tabulating Machine Company, which would later become the International Business Machines Company (IBM)—the company that developed many of the technologies underlying smart cities today.

Finally, since no data collection and analysis methods are immune from error, there have been recurring concerns about the reliability of data reporting. Throughout the history of the census, there have been doubts and disputes about the accuracy of census results, attributed to personal or bureaucratic data manipulation. In fact, even at the time of the first census, George Washington and Thomas Jefferson expressed skepticism about the accuracy of the results. Washington believed residents “did not want to be counted for religious reasons or feared that the census was intended as a foundation for a tax,” while Jefferson estimated that the census population count of just under 4 million had missed around 100,000 residents.<sup>7</sup> These doubts were likely colored by the fact that “political observers equated rapid population growth with economic and political strength, and both men were eager to showcase the new nation’s growing prosperity.”<sup>8</sup> In the 1870 census, population reports reflected drastically lower growth rates than expected. Under pressure from the press, President Ulysses S. Grant ordered census recounts in Philadelphia, New York, and Indianapolis, where estimates suggested the most severe undercounting. Yet recent analysis suggests that retroactive adjustments to census counts may have overcorrected for undercounts, perhaps masking the toll of the Civil War.<sup>9</sup> To further complicate the picture, the public may have given inaccurate information, since “age heaping”—an error that occurs when an individual misstates his or her true age by rounding it to a preferred

age—was common in nineteenth-century censuses and is readily discernible in plots of the population’s age distribution.”<sup>10</sup> These early accounts demonstrate the potential unreliability of data sources and reveal the public’s concerns about privacy, accuracy, and the intentions of government agencies collecting data.

The census is only one example of a governmental institution increasingly relying on rapid and widespread data collection. We can see similar patterns across sectors, such as in education, with the National Center for Education Statistics, which collects, analyzes, and reports data in order to study the condition of education in the United States; in health, with the National Center for Health Statistics, which monitors the nation’s health and provides statistical information to guide governmental policies; and in law enforcement, with COMPSTAT, a technological system and management strategy that uses data analytics to help American police agencies predict and prevent crime.<sup>11</sup> The development of new technologies has increased our capacity to process data and has revolutionized the way we collect and understand data. However, at the same time, the challenges of early data collection resonate with anxieties bound up in today’s data collection practices. The increasing range and scale of data means that the complexity of data collection and analysis has also increased, leaving the public disempowered: power holders collect and control data, while citizens affected by their decisions have little recourse for understanding or challenging these practices.

### **1.1.2 From Planned Data Collection to Ubiquitous Data Collection**

Data collection has shifted away from practices that require large organizations to gather information manually toward systems and technologies that allow and in many cases compel individuals to share information about themselves voluntarily. Through the global spread of wireless network infrastructures that enable individuals to connect to the Internet via mobile devices, data collection has become ubiquitous.<sup>12</sup> Wireless networks not

only enable individuals to be online more often and from more locations: they have also succeeded in bringing individuals on the other side of the digital divide online for the first time. Smartphones, with their built-in microphones, cameras, and GPS are built for data collection and can be used to quantify personal habits. Today, every person who owns a smartphone is constantly collecting and transmitting data over wireless networks.

Many of the largest companies have developed business models that rely on this phenomenon: they depend on generating and collecting data while giving their products away for free. For instance, Google provides a range of services free of charge because the personal data users provide is more profitable than charging for services. Essentially, data has become a more valuable commodity than cash for some businesses—ubiquitous data collection has become a big business.

Whereas governments and institutions once actively collected data from individuals, today, individuals themselves are collecting and transmitting the data, whether or not they are aware of it. Previously, in other words, individuals knew exactly what information they were reporting; yet a key characteristic of ubiquitous data collection practices today is that data is opportunistically and passively collected, whether or not the collectors know—or inform the users—what they are going to do with it.<sup>13</sup> Meanwhile, the new “Internet of things” means that there are many more objects than humans connected to the Internet. Individuals’ devices are also connecting, and everything they do is being measured and transmitted. Most individuals do not understand these processes; they need data literacy in order to understand what data they are transmitting and what is happening with their data.

## 1.2 The Impact of Top-Down Data Management

As data collection has escalated and data has acquired greater value, it has been increasingly used to inform decision-making. We can see this across large organizational settings, from

governments to corporations. This is an evolution that has played out dramatically in the management of cities. Over the years, city governments and their commercial partners have experimented with new waves of data-centered urban models and technologies to improve the efficiency of services in cities and to address the issues that characterize city life, with varying impact on the residents who inhabit them. The promise underlying the use of data in urban planning is the very same thing that makes it potentially problematic: predictive modeling can be used to make assumptions; it is possible to predict the probability of an outcome based on an existing related data set.

### **1.2.1 Data Modeling from Systems Theory to the Smart City**

The systems theory of planning, in which data is used to calibrate self-regulating systems, arose in the 1960s as a response to the growing belief in the inter-relatedness of diverse urban systems and the growing complexity of transportation planning. Systems theory places an emphasis on activity, dynamism, and change, relying on an ongoing process of monitoring, analyzing, and intervening in situations.<sup>14</sup> The movement became more widespread in city planning due to the development of computers capable of handling complex data, and a generation of planners embraced the optimism of the “abstract, highly technical (and frankly abstruse) language of systems theory, with its talk of mathematical modelling, ‘optimisation’ and so on.”<sup>15</sup> This school of thought had its origins in cybernetics, the theory that statistical mechanics can be used to predict the future of a system and that precise modeling of systemic relationships can be used to optimize performance. Systems theory was widely espoused in urban planning literature such as McLoughlin’s 1969 *Urban and Regional Planning: A Systems Approach* and Jay Forrester’s 1969 *Urban Dynamics*, and it was also put in practice to different extents in towns and cities across the United States.

Although systems theory fell out of favor in urban planning because the technology of the 1960s and 70s proved incapable of handling the complexities of urban systems,

numerous corporations continued to develop more sophisticated and automated methods of data modeling that were used in-house to increase efficiency and productivity. Thus, corporations such as IBM, Siemens, and Cisco, which had successfully implemented predictive models on a corporate scale within their own businesses, took a leading role in initiating the resurgence of predictive modeling in urban planning in the early 2000s with the concept of the “ubiquitous city” or “smart city.”<sup>16</sup> For these companies, the smart city was an opportunity to market their technologies as municipal services, thereby developing profitable long-term relationships with city governments.

The smart city is a futuristic, utopian city in which technology and data are optimized to solve the inefficiencies of city management and ease the everyday frictions of city life. In such a city, coordinated traffic-control systems eliminate gridlock, efficiency-optimized systems control electricity flow, and sophisticated monitoring minimizes the impact of natural disasters. The idea of the smart city closely resembles the dream of systems theory advocates in the 1970s. The companies at the forefront of the smart cities movement believe that improved methodology and advanced technology, better able to respond to complex input, will allow a new generation of city planners to realize their utopian vision while avoiding the issues that hindered systems theorists in the past.

The original vision of the smart city was a completely reimagined city, accomplished by constructing experimental utopian cities from the ground up with a new kind of infrastructure built in. New Songdo, South Korea, was an ambitious prototype, begun in 2005, exemplifying this vision. New Songdo was imagined as an energy-efficient “international business district” where all major information systems shared data, connected by a digital infrastructure built into houses, streets, and office buildings.<sup>17</sup> In the early days of its construction, John Kim, vice president for strategy at New Songdo City Development, described it this way: “The same key can be used to get on the subway, pay a parking meter, see a movie, borrow a free public bicycle and so on. It’ll be anonymous, won’t be linked

to your identity, and if lost you can quickly cancel the card and reset your door lock.... [Residents will enjoy] full videoconferencing calls between neighbors, video on demand and wireless access to their digital content and property from anywhere in Songdo.”<sup>18</sup> These services were all to be provided by Cisco, which developed an exclusive partnership with New Songdo to provide digital infrastructure and services. For Cisco, the partnership represented the kick-off of a new industry, whereby the company would offer all urban necessities as a single internet-based utility. Today, New Songdo is only half-finished and is a financial and technological disappointment. Townsend points out that given the innovations in the years since construction began on New Songdo, building a city around RFID seems anachronistic.<sup>19</sup>

Today, the term “smart city” more aptly refers not to new, specially designed cities, but rather to the practice of retrofitting existing cities to make them “smarter,” a concept popularized by IBM’s “Smarter Planet” campaign, which was originally launched in 2008. In a red paper published in 2011, IBM defines a smarter city as “one that makes optimal use of all the interconnected information available today to better understand and control its operations and optimize the use of limited resources.”<sup>20</sup> IBM partnered with Portland, Oregon in 2009 to model the relationships among the city’s economy, housing, education, public safety, transportation, healthcare, government services, and utilities in order to simulate how city systems work together and to suggest opportunities for innovation. The goals of the project epitomized systems theory: to develop a model to improve long-range planning and help the municipal government reduce negative consequences of their services and to uncover opportunities for efficiency.<sup>21</sup> In 2011, IBM partnered with Rio de Janeiro’s mayor, Eduardo Paes, to create a sophisticated Operations Center that predicted rainfall and flooding across the city, mapped car accidents and power failures, and integrated information coming in from phone, radio, e-mail, and text message reports so that government services could respond more quickly to events and access real-time information

while in the field.<sup>22</sup> The solutions for Rio were an effort to augment flawed infrastructures incapable of responding to crime and disaster. The dream of the smart city is still pervasive today. A recent article in *China Daily* describes China's current preparation for a "smart city construction boom," initiated in November 2012, which will implement dramatic changes in 193 trial cities, in fields ranging from transportation to the financial sector. Both Beijing and Shanghai will focus on developing a network to "facilitate transportation, tele-medicine, and smart homes," while Shanghai will specifically focus on "developing wireless broadband technology and boosting the application of intelligent technology."<sup>23</sup> These efforts to smarten up cities in China, as well as the models that preceded them, rely on predictive modeling to radically change existing infrastructures.

### **1.2.2 Problems of Data Modeling**

Although data modeling has in many instances succeeded in improving efficiency, urban planners have repeatedly failed to meet their objectives as they put data analytics into practice in city management. Predictive modeling's first great flaw is that it rarely accounts for human error; data is seen as being entirely objective, not vulnerable to mistakes in collection or interpretation. Second, modeling tends to oversimplify complex factors and point toward a single solution. Finally, the rhetoric of predictive modeling was developed around the singular goal of improving efficiency, a motivation that was heavily influenced by corporations and thus isn't always in the best interest of city residents. There often emerges, in other words, a predictable conflict between profit-driven models and the needs of citizens, needs which do not always fit neatly into monetizable problem-solution models.

First, there is minimal acknowledgement of the possibility of human error or misuse in data modeling: data is often assumed to be more objective than human perception. Yet there are ample opportunities for human error in the collection or analysis of data due to inaccurate or biased methods of data collection, interpretation, and application. "Numbers

can't speak for themselves, and data sets—no matter their scale—are still objects of human design.... [T]here is a problematic belief that bigger data is always better data and that correlation is as good as causation.”<sup>24</sup> Predictive city modeling developed from a desire to improve services. Yet its overreliance on opaque technologies, without a corresponding effort to educate the public or even the policy makers responsible for making data-based decisions, resulted in misuse by decision makers. This is the peril of an overreliance on data; when data modeling is used without taking into consideration existing local knowledge, there can be huge blind spots, intentional or not. Exacerbating this problem is the fact that relying on conclusions drawn from data, using abstruse methods that seem authoritative, makes policy harder for opponents to critique.

Adam Greenfield and Anthony Townsend both point to one of the most devastating failures of data modeling in the 1970s, the fires that resulted from the RAND Corporation’s work for the New York Fire Department, as recounted in Joe Flood’s 2010 *The Fires*.<sup>25</sup> New York City mayor John Lindsay and fire commissioner John O’Hagan sought the advice of the RAND Corporation, a think tank, to use statistics and systems analysis to help optimize performance of the fire department. RAND conducted a series of studies between 1973 and 1975 on FDNY response-time to determine the optimal distribution of fire stations. Due to inaccurate reporting of data by fire captains and simplified formulas that failed to account for fire station capacity and the effect of traffic on response time, RAND recommended the closure of numerous fire stations concentrated in poor areas of the city. When these closures were implemented, many of the remaining fire companies were overwhelmed and numerous fires in the Bronx and other areas displaced more than a half-million residents.<sup>26</sup> In this instance, data modeling was used to provide “scientific” evidence that confirmed Mayor John Lindsay’s existing biases against stations where union leaders were based. This failure came at the expense of many of the city’s most underprivileged communities.

A more recent example that illustrates skewed data modeling is Google's 2012 Flu Trends predictions, which vastly overestimated flu rates based on the prevalence of flu-related Internet searches.<sup>27</sup> Google, whose prediction was twice as high as the Center for Disease Control's prediction the same year, discovered that the prevalence of Internet searches for flu-related terms does not correlate directly with flu rates. The 2012 snafu did not have fatal consequences, but it serves as a reminder that predictive probabilities are not always accurate.

Second, the logic of the smart city relies on the assumption that data will intrinsically reveal solutions, rather than just informing them. Implicit in the rhetoric of the smart city is an “unreconstructed logical positivism” which suggests the world is perfectly knowable and its relations can reliably be encoded in data sets without distortion.<sup>28</sup> “As applied to the affairs of cities, it is effectively an argument there is one and only one universal and transcendently correct solution to each identified individual or collective human need; that this solution can be arrived at algorithmically, via the operations of a technical system furnished with the proper inputs; and that this solution is something which can be encoded in public policy, again without distortion.”<sup>29</sup> This is the belief that creating a better model—one that accounts for complex inputs—will fix the errors encountered in the past. Yet there will always be gaps or blind spots in data, and by indiscriminately following models without fully realizing their limitations, governments hazardously sidestep the democratic process. And since plans for smart cities so often suggest sweeping infrastructural change, the stakes and the risks are high.

Finally, since corporations shaped the agenda of the smart city, the primary professed goal is a corporate agenda: to improve efficiency. Yet policies that emphasize efficiency at the sacrifice of other values are often not in the best interest of citizens. As Adam Greenfield argues, the smart city rhetoric is a form of authoritarianism that takes everyday decision-making power from citizens and places it in the hands of corporations.

He argues that part of the smart city strategy is to change the behavior of residents to help improve the efficiency of the new model—a form of collusion that encroaches on autonomy and self-determination.<sup>30</sup> In other words, instead of creating technologies to better serve people, smart cities demand that residents change their lifestyles to serve technology and corporate interests.

These issues all have a profound impact on individual residents. Because data collection is ubiquitous in everyday technologies, individuals themselves are being modeled. Governments model their statistical data; corporations model their consumer data. In this scenario, individuals have become the product. They are not consulted or considered in top-down decisions that deeply impact them. Most individuals do not know what predictive data modeling is, let alone have the knowledge to evaluate and point out the shortcomings of the complex and opaque urban models implemented in their cities. Clearly, the public needs better data literacy in order to understand and question these decisions and processes.

## **1.3 A Counter Move: Open Data and Bottom-Up Data Innovation**

The open data movement attempts to correct this problematic asymmetry between the impact of data on individuals and their ability to conceptualize data by making local, regional, and national data, especially publicly acquired data, available to the public in a standardized form that allows for direct manipulation.<sup>31</sup> In this movement, hackers, journalists, and activist organizations are working with data to draw attention to important issues and to build bottom-up services and technologies to respond to issues in their communities. Embedded within the growing culture of open data is the understanding that data is a publicly owned asset.

### **1.3.1 Institutions Move Toward Open Data**

An increasing number of countries across the globe are implementing open data policies. Open government data comprises an important component of the Open Government Partnership, an initiative launched in September 2011 that includes 55 countries. Online data portals such as data.gov in the United States, launched in May 2009, and data.gov.uk in the UK, launched in January 2010, have served as models for efforts in numerous other countries. Although there are no recognized standards that define what open data really entails, the Open Data Index,<sup>32</sup> an effort to measure and classify open data offerings across the globe, has defined a set of criteria for classifying open data initiatives that gives us a sense of current expectations. The Index looks at the availability of data such as transportation timetables, government spending, and legislation and asks a series of questions to determine how public and accessible this data is online: Does the data exist? Is it digital? Is it publicly available? Is it free of charge? Is it online? Is it machine-readable? Is it available in bulk? Is it openly licensed? Is it up-to-date? According to the Index, the UK, US, and Denmark rank highest in openness, but most of the 70 countries surveyed meet at least some of the criteria.

The underlying idea of the movement is that public data should be available online in order to increase transparency and accountability and promote public participation in decision-making and social innovation. In President Obama's words: "One of the things we're doing to fuel more private-sector innovation and discovery is to make vast amounts of America's data open and easy to access for the first time in history. And talented entrepreneurs are doing some pretty amazing things with it."<sup>33</sup> Of course, governments still withhold much of the data they collect; in light of the NSA's revelations about PRISM, for example, Obama's words must be taken with a grain of salt.

Democratic governments have been buying into the open data movement for two reasons: first, it is in their best interest to appear receptive to the input of everyday citizens;

second, they are responding to pressure from civil society over the last several years. In Iceland, for instance, the open data movement gained traction after the collapse of the financial system in 2008, when negligence by government and regulators was blamed. After the collapse, there was an increased demand for transparency and access to public data.<sup>34</sup> To restore public confidence, the Icelandic government made open government a high priority by implementing policies such as opening governmental data and even crowdsourcing a constitutional bill, which ultimately failed but set a precedent for a new participatory decision making process.<sup>35</sup>

Access to open government data can be seen as an extension of citizens' right to information, the notion that the public has the right to access information held by government on request and that governments have an obligation to actively disseminate information to the public on matters of public interest. Although the right to information has long been considered a basic right and has informed legislation in many countries, open government data is still seen as a privilege, and there are not yet established standards about how data should be shared.<sup>36</sup> It is very much an evolving practice that is currently in a critical moment of definition.

The open data movement has propagated a shifting notion of who the users of data are. In the long history of data, citizens were always considered to be end-users who provided their data to the collector and then interfaced with the end-products of data-driven government innovations. In this new vision, government concedes that citizens can best define and resolve the problems that plague their own communities—implying that communities should take the data provided and use it to address their needs.

### **1.3.2 Grassroots Data Collection and Visualization**

Many organizations and individuals have taken advantage of open data resources. There is a growing group of not-for-profits that use open data to address issues in local communities,

often with the collaboration or sponsorship of local governments, innovating in ways that government cannot. For example, Code for America’s 2013 National Day of Civic Hacking encouraged programmers across the country to partner with organizations and city governments to use public data for local problem-solving. Data Kind, a not-for-profit in New York City, aims to connect volunteer data scientists and developers with social organizations that want to use insights from data to serve their communities but lack the expertise or resources to do so. These organizations aim not only to solve specific problems but also to spread awareness and to create a community interested in using technology and data for civic innovation.

The result of such efforts are applications like Code for America’s “Adopt-a-Hydrant,” which called on Boston residents to help dig out fire hydrants during snow storms, “Discover BPS,” which helped Boston residents navigate complex requirements for choosing a public school, and “Where’s My School Bus?” which enabled parents to track their child’s bus in real-time. These interventions set a precedent for using technology to enable citizens and government to work together to develop flexible bottom-up responses to city issues using open data.

In addition to using open government data, many organizations have also created applications that enable everyday citizens to provide information reflecting their local knowledge and filling in gaps in government data. For instance, Ushahidi is a participatory mapping platform that was originally developed in 2008 to allow Kenyans to submit reports of violent incidents, using the Internet and mobile phones.<sup>37</sup> The platform is now used by organizations around the world to crowdsource information. Another example is SeeClickFix, an online 311 platform that allows citizens to report neighborhood issues. These applications rely on everyday citizens to identify the issues that concern them and to help collect and visualize data.

### **1.3.3 Limitations of the Open Data and Grassroots Data Movements**

Despite the promise of open government and grassroots data movements, most individuals are afforded only minimal participation in data practices. While in a growing number of instances everyday citizens have been encouraged to take a more active and deliberate role in data collection, the civic hackers creating these applications and visualizations, and therefore setting their agenda, are expert data scientists, programmers, and designers.

Although data has become ubiquitous and has a profound impact on individuals' lives, the public's understanding of data has not grown proportionally. The open data movement suggests that individuals and communities should use government data to problem solve and become more engaged with policy, yet it does not equip them with the resources to do so. Instead, governmental agencies appear to want to transfer the responsibility for data interpretation to citizens who do not possess the necessary skills for collecting or processing data.

Most individuals are still excluded; even in cities and countries that are at the forefront of the open data movement, there is a data divide. We can dissect this problem using the lens of the four critiques of open data set forth by Rob Kitchin: "open data lacks a sustainable financial model; promotes a politics of the benign and empowers the empowered; lacks utility and usability; and facilitates the neoliberalisation and marketisation of public services."<sup>38</sup> I want to focus on two of the issues that Kitchin identifies: that open data empowers the empowered, and that it lacks utility and usability.

First, I will address the question of empowerment. Most of the individuals who are currently taking advantage of open data resources belong to an elite community of open data and transparency activists and civic hackers—those who were lobbying for access in the first place, and specialists in data science, who built businesses around open data. These practitioners often do innovative work with widespread benefits, but they do not always represent a larger public because data mining is not a practice accessible to everyday citizens.

These practices do, in a sense, represent bottom-up innovation, since individuals and small organizations are working within alternative infrastructures independent of government or industry. However, when independent organizations define problems or propose solutions, they do not necessarily consult or represent the communities they are attempting to assist.

Since data requires access to education, infrastructure, and technology, it favors people who have more resources over those without means. Solly Benjamin and his colleagues discovered one example of this unequal advantage while studying the impact of the digitization of land records in Bangalore.

Their findings were that newly available access to land ownership and title information in Bangalore was primarily being put to use by middle and upper income people and by corporations to gain ownership of land from the marginalized and the poor. The newly digitized and openly accessible data allowed the well-to-do to take the information provided and use that as the basis for instructions to land surveyors and lawyers...to challenge titles, exploit gaps in titles, take advantage of mistakes in documentation, [and] identify opportunities and targets for bribery.<sup>39</sup>

In addition, applications that seek citizen participation are less likely to capture the opinions of the disadvantaged, who may lack access to the necessary technology. Kate Crawford points out that, “if cities begin to rely on data that only comes from citizens with smartphones, it’s a self-selecting sample—it will necessarily have less data from those neighborhoods with fewer smartphone owners, which typically include older and less affluent populations.”<sup>40</sup> For instance, applications like SeeClickFix and the Boston Office of New Urban Mechanics’ Street Bump app, which collects smartphone data from drivers going over potholes, are both affected by this issue.

Second, open data lacks utility and usability. Open data movements require a genuine commitment from governments to maintain data portals in order to provide complete and up-to-date information. They must also commit to publishing data unconditionally, despite the fact that the results might be critical of government policies.

In 2011, Kenya’s President Mwai Kibaki initiated the Open Data Initiative to increase

government transparency. There was a great amount of excitement about the launch, because Kenya was only the second African country to open its data. Yet as of 2013, the portal has not been updated in nearly a year, and traffic is stagnant. The government has been reluctant to release its data, and most Kenyans either do not know how to use the site or are not aware it exists.<sup>41</sup> To truly increase accessibility and utility, open data initiatives must provide data that is complete, up-to-date, and useful to the public.

Governments have been struggling to meet this last criterion, which relates to the form data takes and how easy it is for individuals to interpret. Data is difficult to understand; it is rarely possible to simply look at a data set and draw conclusions from it. We require aids for comparing data in a way we can understand. Raw data sets, in the form of spreadsheets or lists, are useless to the general public since analyzing and drawing conclusions from raw data requires access to sufficient education and technology.

Compounding this difficulty is the issue that standard methods of communicating data to the public—namely data visualizations—are also problematic. Visualizations are often used as a gesture toward accessibility rather than as an actual window into content. They are not always accurate, easy to interpret, or neutral. The purpose of data visualization is to “discover the structure of a (typically large) data set. This structure is not known *a priori*; a visualisation is successful if it reveals this structure.”<sup>42</sup> Yet every data set contains multiple truths, and it is impossible to capture every truth in a visual representation. Visualization is therefore an act of interpretation, not inherently one of communication. We can’t rely on institutions to process and present data sets for us in an objective manner, because beliefs and biases are encoded into visualizations. Visualization is used just as often to obscure information as to reveal it.

## 1.4 The Public Needs Better Data Literacy

Individuals are increasingly impacted by data wielded by government and corporate institutions. Meanwhile, the rising culture of open data and civic hacking has made data more visible, more accessible, and more malleable. Despite data's problems and ambiguities, the ability to collect and analyze data is a powerful competency that allows individuals to navigate our data-rich society and innovate for social good. The public has a new opportunity to participate in the data dialogue, but they lack the resources and expertise to do so. The public needs better data literacy—and I believe this is especially important for young people so they learn to participate in an increasingly data-rich society. By supporting young people in developing data literacy, we are giving the data scientists of the future the tools to think critically and ethically about data.

In this Chapter, I illustrated the societal factors that have made data literacy so important today and described why we have a new opportunity for supporting data literacy among the public. In the following chapters, I will explore what it means to be data literate, how initiatives in different learning environments are supporting the development of data literacy, and what strategies we can adopt to advance this field in the future.

### NOTES

- 1 H. C. Darby, *Domesday England* (Cambridge University Press, 1986), 336.
- 2 Census History Staff US Census Bureau, “1790 Overview - History - U.S. Census Bureau,” accessed December 11, 2013, [http://www.census.gov/history/www/through\\_the\\_decades/overview/1790.html](http://www.census.gov/history/www/through_the_decades/overview/1790.html).
- 3 Census History Staff US Census Bureau, “Counting the Population - History - U.S. Census Bureau,” accessed December 11, 2013, [http://www.census.gov/history/www/innovations/data\\_collection/counting\\_the\\_population.html](http://www.census.gov/history/www/innovations/data_collection/counting_the_population.html).

- 4 In 1940, with the introduction of statistical sampling, there was a shift to asking only a random sample of the population to provide additional information and using their results to extrapolate information about the rest of the population, marking an evolution toward the predictive data modeling that is pervasive today.
- 5 Anthony M. Townsend, *Smart Cities : Big Data, Civic Hackers, and the Quest for a New Utopia* (New York: W.W. Norton & Company, Inc., 2013), 59.
- 6 Ibid., 60.
- 7 J. David Hacker, “New Estimates of Census Coverage in the United States, 1850-1930. (cover Story),” *Social Science History* 37, no. 1 (Spring 2013): 75, doi:10.1215/01455532-1958172.
- 8 Ibid.
- 9 Ibid., 95.
- 10 Ibid., 83.
- 11 “Standards for Education Data Collection and Reporting (SEDCAR),” December 31, 1991, <http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=92022>; “About NCHS - NCHS Fact Sheets - Overview,” accessed May 5, 2014, [http://www.cdc.gov/nchs/data/factsheets/factsheet\\_overview.htm](http://www.cdc.gov/nchs/data/factsheets/factsheet_overview.htm); James J. Willis, Stephen D. Mastrofski, and David Weisburd, “Making Sense of COMPSTAT: A Theory-Based Analysis of Organizational Change in Three Police Departments,” *Law & Society Review* 41, no. 1 (March 1, 2007): 147–88, doi:10.1111/j.1540-5893.2007.00294.x.
- 12 Townsend, *Smart Cities*.
- 13 Kate Crawford et al., *Big Data, Communities and Ethical Resilience: A Framework for Action*, Community Resilience through Big Data and Technology (Bellagio Center: Rockefeller Foundation, 2013), 2.
- 14 Nigel Taylor, *Urban Planning Theory Since 1945* (SAGE, 1998), 63.
- 15 Ibid., 64.
- 16 Townsend, *Smart Cities*.
- 17 Pamela Licalzi O’Connell, “Korea’s High-Tech Utopia, Where Everything Is Observed,” *New York Times* 5 (2005), <http://tofugalaxy.com/wp-content/uploads/2010/11/nytimes.pdf>.
- 18 Ibid.
- 19 Townsend, *Smart Cities*, 26.
- 20 Michael Kehoe et al., “Smarter Cities Series: A Foundation for Understanding IBM Smarter Cities,” *An IBM Redguide Publication*, 2011, <http://www.redbooks.ibm.com/redpapers/pdfs/redp4733.pdf>.

- 21 “IBM and City of Portland Collaborate to Build a Smarter City,” CTB10, August 9, 2011, <http://www-03.ibm.com/press/us/en/pressrelease/35206.wss>.
- 22 Natasha Singer, “I.B.M. Takes ‘Smarter Cities’ Concept to Rio de Janeiro,” *The New York Times*, March 3, 2012, sec. Business Day, <http://www.nytimes.com/2012/03/04/business/ibm-takes-smarter-cities-concept-to-rio-de-janeiro.html>.
- 23 “China Prepares for Smart City Construction Boom,” accessed December 13, 2013, [http://www.chinadaily.com.cn/business/2013-10/16/content\\_17035874.htm](http://www.chinadaily.com.cn/business/2013-10/16/content_17035874.htm).
- 24 Kate Crawford, “Think Again: Big Data,” *Foreign Policy*, May 10, 2013, [http://www.foreignpolicy.com/articles/2013/05/09/think\\_again\\_big\\_data](http://www.foreignpolicy.com/articles/2013/05/09/think_again_big_data).
- 25 Adam Greenfield and Nurri Kim, *Against the Smart City (The City Is Here for You to Use)*, 1.3 edition (Do projects, 2013), sec. 4; Townsend, *Smart Cities*, 80.
- 26 Joe Flood, *The Fires: How a Computer Formula Burned down New York City—and Determined the Future of American Cities* (New York: Riverhead Books, 2010), 161.
- 27 Declan Butler, “When Google Got Flu Wrong,” *Nature* 494, no. 7436 (February 13, 2013): 155–56, doi:10.1038/494155a.
- 28 Greenfield and Kim, *Against the Smart City (The City Is Here for You to Use)*, sec. 4.
- 29 Ibid.
- 30 Ibid., sec. 11.
- 31 Michael B. Gurstein, “Open Data: Empowering the Empowered or Effective Data Use for Everyone?,” *First Monday* 16, no. 2 (January 23, 2011), <http://journals.uic.edu/ojs/index.php/fm/article/view/3316>.
- 32 <http://index.okfn.org>.
- 33 Crawford, “Think Again.”
- 34 “Iceland: From the Financial Crisis to Open Data,” *Open Knowledge Foundation Blog*, accessed December 12, 2013, <http://blog.okfn.org/2010/03/26/iceland-from-the-financial-crisis-to-open-data/>.
- 35 “Crowdsourcing Iceland’s Constitution,” *IHT Rendezvous*, accessed December 12, 2013, <http://rendezvous.blogs.nytimes.com/2012/10/24/crowdsourcing-icelands-constitution/>.
- 36 Janssen, *Open Government Data*.
- 37 <http://ushahidi.com>.
- 38 Rob Kitchin, “Four Critiques of Open Data Initiatives,” *Impact of Social Sciences*, November 27, 2013, <http://blogs.lse.ac.uk/impactofsocialsciences/2013/11/27/four-critiques-of-open-data-initiatives/>.

- 39 Gurstein, “Open Data.”
- 40 Crawford, “Think Again.”
- 41 Greg Brown, “Why Kenya’s Open Data Portal Is Failing - And Why It Can Still Succeed,” *Sunlight Foundation*, accessed December 7, 2013, <http://sunlightfoundation.com/blog/2013/09/23/why-kenyas-open-data-portal-is-failing-and-why-it-can-still-succeed/>.
- 42 Lev Manovich, “What Is Visualisation?,” *Visual Studies* 26, no. 1 (March 2011): 36–49, doi:10.1080/1472586X.2011.548488.



## CHAPTER 2

# Data Literacy: Definition and Methods

As I illustrated in Chapter 1, the public needs better data literacy. This means being able to understand and work with different types of data—from quantitative data to qualitative data, and from very large datasets to very small ones. The notion of being “literate” with data is expansive, because the term “literacy” has come to represent such a wide spectrum of knowledge and skills, far beyond its original definition of reading and writing. While “literacy” emerged as “a new word invented to express the achievement and possession of what were increasingly seen as general and necessary skills,”<sup>1</sup> it has been liberally appended to various concepts to convey that it is crucial for the public to understand them.<sup>2</sup> Data is the newest of several new media to which the idea of literacy has been applied in this way, signaling aspirations that its pursuit will produce empowered individuals, better-informed communities, and a more skilled workforce.

Although data literacy is useful in many contexts and for many audiences, most of the research to define data literacy competencies and develop initiatives to support them is focused on higher education—data literacy initiatives that train researchers and professionals to use data in their fields. In this chapter, I propose a definition of data literacy that is applicable to the general public, including youth, and I describe a range of methods for supporting youth data literacy.

This chapter is organized into two sections. In the first section, “What is data literacy?” I first propose a definition of “data literacy” composed of a set of core competencies. Next, I draw from the literature on traditional and new media literacies in order to situate the concept of data literacy in terms of its societal impact, objectives, disciplinary origins, and theoretical foundations. Finally, I describe the personal and societal benefits of data literacy. In the second section, “How Can We Support Data Literacy?” I describe a range of methods for helping youth develop data literacy. I draw from examples of educational technologies that aid in data exploration, out-of-school models for learning with data, and models for incorporating data literacy into classroom learning.

## 2.1 What is Data Literacy?

### 2.1.1 Data Literacy Competencies

Since data is used differently in various domains, researchers have proposed multiple possible definitions of the competencies needed to be data literate. These definitions differ in terms of the skills they emphasize, the level of technical proficiency they call for, and the methods and technologies they specify.

Researchers in scientific fields, for instance, focus on the highly technical skills related to collecting and managing quantitative data in order to conduct scientific research. A study by Carlson et al. assessed the data research and management skills used by a group of faculty in science and engineering fields at Purdue University and the University of Illinois at Urbana-Champaign in order to identify twelve key areas of data literacy competency: introduction to databases and data formats; discovery and acquisition of data; data management and organization; data conversion and interoperability; quality assurance; metadata; data curation and re-use; cultures of practice; data preservation; data analysis; data visualization; and ethics.<sup>3</sup>

Researchers in information science, on the other hand, emphasize the mechanics of finding and managing data. For instance, Prado and Marzal offer a framework of core competencies for data literacy instruction within information literacy programs in libraries:<sup>4</sup>

1. Understanding data: What is data and what is data's role in society?
2. Finding and/or obtaining data: What are possible data sources, how can they be evaluated, and how can new data be obtained?
3. Reading, interpreting and evaluating data: How can data be read, interpreted, and evaluated?
4. Managing data: How should data and metadata be collected and managed?
5. Using data: How should data be handled, how are elements produced for data synthesis, and what are the ethical considerations?

Although most discussions of data literacy focus on higher education, there are also definitions of data literacy intended for the general public. For instance, the School of Data, a collaboration between the Open Knowledge Foundation and Peer to Peer University, provides data literacy training for the public.<sup>5</sup> Their “Data Fundamentals” course covers six topics: what is data?; finding data; sorting and filtering data; analyzing data; creating data visualizations; and telling stories using data.

I believe it is important to embrace a broad definition of data literacy that accounts for this diversity. Drawing from these definitions and my broader research, I propose a working definition: data literacy is the ability to understand, find, collect, interpret, visualize, and support arguments using quantitative and qualitative data.

**1. Understanding Data:** Data literacy requires an understanding of what data is, what types of data exist, and how data is generated. A data literate individual must also have knowledge of the role and impact of data in society across different contexts and the ethical implications of using data.

**2. Finding Data:** Data literacy requires the skills necessary to ask questions that can be researched using data, find relevant data sources, compare and evaluate sources, and check for bias and inaccuracy.

**3. Collecting Data:** Data literacy encompasses the ability to collect both qualitative and quantitative data, including methods such as conducting interviews, creating surveys, making observations, and taking measurements.

**4. Interpreting Data:** Data literacy requires the ability to prepare data for analysis, critically analyze data in a range of formats, and develop inferences.

**5. Visualizing Data:** Data literacy includes the ability to communicate data using a range of visual representation methods such as tables, graphs, and maps.

**6. Supporting Arguments Using Data:** Data literacy includes the ability to use data as evidence to support arguments and tell stories, taking into consideration the larger cultural, social, and political implications of data-derived insights. Ultimately, data literate individuals should be able to use data to solve problems and communicate their solutions.

One aspect of this definition that sets it apart from others is that I consider the ability to think across methods to be a core component of data literacy. Current approaches to the study of data tend to be determined along methodological boundaries, privileging either quantitative data analysis or qualitative investigation. Quantitative analysis is more likely to be taught in a mathematics or science context, while qualitative research might be taught in a social science or journalism context. I believe that this division is limiting

in terms of learning about larger patterns and about individuals. Analyzing quantitative data can illuminate hidden patterns and reveal insight into complex social phenomena. But quantitative analysis requires individuals to be grouped into categories, either by location or demographic attributes; individual stories and motivations get lost in those aggregations. On the other hand, a qualitative analysis that focuses solely on personal accounts or narratives, without knowing how typical they are, may be misleading and does not illuminate the larger picture. By analyzing quantitative data alongside qualitative data, the insights from quantitative data can be contextualized by human stories. This approach can produce a more balanced and grounded perspective. Exposing individuals to both types of inquiry enables the exploration of an issue through multiple lenses.

These competencies are intended to serve as general learning objectives for initiatives that aim to support youth data literacy. Yet every program is different and will need to customize its approach to focus on the competencies appropriate for its own mission and audience. These skills have different levels of complexity, so a program supporting a basic level of data literacy might skip lessons in manipulating raw data, focusing instead on analyzing preexisting data visualizations. Building data literacy is not about mastering any single skill or technology platform. It is about building the understanding and competencies that will best empower individuals in their interactions with data.

### **2.1.2 Situating Data Literacy**

Data literacy is related to several other types of literacy—information literacy, media literacy, quantitative literacy (or numeracy), and computational thinking—that have been widely discussed and researched. In order to better define data literacy’s theoretical approach and what it aims to accomplish, we can situate it among these other literacies and the disciplines from which they are derived.

Data literacy is at the intersection of information and media literacy, which represent two theoretical approaches to the notion of literacy. Information literacy is a pre-Internet era concept that originated in Library and Information Sciences and aimed to prepare individuals to be productive citizens in an “information age.” Information literacy tends to put an emphasis on locating and determining the credibility of information; the concept of information literacy has been extended to apply to navigating information online. A digitally literate individual, then, would be someone who can search for sources, compare them, and assess their credibility.<sup>6</sup> The concept of information literacy has been criticized for its focus on building technical skills that are easy to acquire but quickly become obsolete. Discussions of media literacy, on the other hand, deemphasize skill acquisition and focus instead on ways of supporting media production and a deeper critical understanding of media focused on modes of representation, language, production, and audience.<sup>7</sup> Data literacy is related to information literacy because the skills required to locate and evaluate data, a type of information, are core components. At the same time, data literacy draws on the concept of media literacy insofar as it places an emphasis on understanding the ways data is represented through various forms of media and on locating those representations culturally.

Data literacy is also closely related to quantitative literacy (or numeracy), which is “characterized by the use of simple quantitative tools to deal with complex issues.”<sup>8</sup> Proponents of quantitative literacy argue that strategies for analyzing numeric problems are similar across many contextual domains, and that it is important for individuals to be able to use simple mathematical reasoning and data comparison within their everyday lives.<sup>9</sup> For instance, a numerate individual would be able to calculate how to split a lunch bill, reconcile a bank statement, and read a bus schedule. Like quantitative literacy, data literacy implies the ability to reason critically, statistically, and mathematically. Yet while data literacy supports and draws on these principles, it is a distinct area of knowledge that extends

beyond quantitative literacy by emphasizing fluency with both quantitative and qualitative data. It relies on an exploratory inquiry process of data collection and analysis that resembles scientific inquiry more than mathematical problem solving.<sup>10</sup> Data literacy also includes an understanding of the logic and ethics involved in making decisions based on data, as well as a focus on communicating data through design.

Finally, data literacy can be thought of as a subset of computational thinking, which is “an approach to solving problems, designing systems and understanding human behavior that draws on concepts fundamental to computing.”<sup>11</sup> Computational thinking includes seeking algorithmic approaches to problems, an ability to move between different levels of abstraction, and a focus on modeling as a way to identify relationships.<sup>12</sup> Computational thinking enables individuals to think the way a computer “thinks” in order to give it instructions through programming. Data literacy focuses specifically on the computational processes related to the collection and use of data, a domain that falls within computational thinking but outside its focus.

### **2.1.3 Benefits of Data Literacy**

We can think of literacy as both an individual good, building individuals’ cognitive skills, and as a societal good, building a more informed citizenry. Livingstone et al. propose that any media or information literacy has three purposes: 1. democracy, participation, and active citizenship; 2. knowledge economy, competitiveness, and choice; and 3. lifelong learning, cultural expression, and personal fulfillment.<sup>13</sup> We can use these principles to examine the civic, economic, and cultural benefits that data literacy initiatives create for both individuals and society.

First, data literacy should support democracy, participation, and active citizenship by equipping learners with skills that can be applied to community problem solving and thereby enabling them to better understand, analyze, and participate in complex social

and political issues. For instance, at a low level of engagement, a data literate individual would be able to read and interpret a data visualization in the news, identify bias, and use that knowledge to inform her opinion. At a high level of engagement, a data literate individual would be able to identify a problem in her city—such as unsafe bike lanes, vandalism, or poor street lighting; look for datasets provided by the city government related to the problem; analyze that data to learn more about the problem and its causes; and put pressure on the government or form strategies for community intervention. Youth data literacy initiatives both in and out of schools should not only teach skills but also serve as an opportunity for participants to critically engage with culturally relevant topics. Whether or not youth are explicitly participating in civic issues in the context of a data literacy program, they will acquire problem-solving tools that will serve as a foundation for future civic engagement. Data can be used to interpret society, highlight inequality, promote tolerance, and resolve conflict. Data literate individuals are more likely to gain informed opinions on current topics, be able to express their opinions, and be able to propose informed solutions; a data literate society would be more capable of critical and sophisticated self-reflection.

There has been significant work framing the literacy discussion as a way to democratize access and participation, focusing on the collective social benefits of literacies and the implementation of policies to spread literacies widely.<sup>14</sup> Many scholars have shifted away from discourse about a “digital divide” in the US to focus instead on the notion of a “participation gap.” While some are still excluded from regular technology usage because of prohibitive costs, Internet access among low-income families has risen dramatically in the last few years.<sup>15</sup> Yet even if underserved youth have access to technology, they are nevertheless less likely than their peers to have the resources and support to participate fully in technology’s educational opportunities. Any digital literacy debate should focus on engaging underserved youth. This is especially true of data literacy, since youth from

wealthier families are more likely to have developed advanced skills, like programming, and basic skills, like map reading, that give them a head start with data literacy. A focus on providing access to data literacy training is critical for closing the literacy gap.

Second, data literacy should support the knowledge economy, competitiveness, and choice by preparing learners with highly marketable skills in the emerging field of data science. A report conducted by the McKinsey Global Institute projects that by 2020 the United States alone will need 140,000 to 190,000 more workers with deep analytical expertise, in addition to 1.5 million managers who are data literate.<sup>16</sup> This growing need reflects the fact that Big Data is changing the way institutions across many sectors make decisions. By supporting young people of diverse backgrounds in developing data literacy, we can not only help them acquire valuable job skills to prepare them for these opportunities but also increase the likelihood that the people who ultimately fill these positions are capable of making informed and ethical decisions with data.

Finally, data literacy should support lifelong learning, cultural expression, and personal fulfillment by enabling learners to pursue creative uses of data. Literacy entails more than the acquisition of skills; a true literacy allows for creative thinking within a pedagogical framework.<sup>17</sup> Once youth understand the conventions of working with data, they are able to not only apply their knowledge to new subjects but also begin developing creative applications for data. For instance, a teenager who is active on Twitter might create an algorithm to analyze the emotions of the Twitter users he follows during a political election, or a child who likes to play video games might learn to develop an online game using data. Youth data literacy programs both in and out of schools should aim to increase participants' interest in interacting with data. They should deeply engage participants in investigating an issue or creating a product, offering the rewards of community support, personal fulfillment, or public recognition, thereby increasing interest and motivation. Through these interactions, as data becomes more legible, participants should become more

likely to seek out data in the future, applying their knowledge about data in new contexts. This not only gives individuals a means for expression but also contributes to an informed and creative society.

## 2.2 How Can We Support Data Literacy?

Behaviorist learning principles, which depend on a teacher imparting instruction and knowledge, are hard-wired into the U.S. educational system. In this approach, the teacher is viewed as an expert who is in control of the learning process, while learners passively obey instructions.<sup>18</sup> Curricula based on behavioral theories are typically designed around specific goals of skills acquisition, a set plan for sequence and content, and direct instruction.<sup>19</sup> Schools typically take a behaviorist approach to curriculum development because such an approach makes it possible to standardize and enforce standards of quality across large organizations.<sup>20</sup> Yet direct instruction in data-related skills and principles is not the best approach to supporting data literacy. Since most youth do not have a predetermined interest in working with data, it is critical for them to experience data's real-world impact at the same time they are building skills.

Instead, we can look to cognitivist or constructivist principles to guide the development of programs that address data literacy. These approaches shift the emphasis from teaching to learning. In his influential essay, "The School and Society," John Dewey argued that learning should be a social and experiential process and that individual students should have a role in determining what to learn based on their own interests.<sup>21</sup> Schools should teach more than knowledge; they should teach life skills. Jean Piaget extended Dewey's ideas with his doctrine of constructivism, arguing that knowledge cannot be transmitted to another person; all learning takes place by discovery. Seymour Papert took this idea a step further with the notion of constructionism, suggesting that learning happens

best when learners actively create meaningful products. In these visions, teachers act as mentors and guides, helping students to explore new topics rather than directly instructing them. Curricula are based on a set of educational experiences that center on creative problem solving, include lessons based on life experiences, champion group learning, and give students input into the curriculum.<sup>22</sup>

Although I believe that some skills-based instruction is necessary to support students in learning complex tasks, skills should not be divorced from their application to real-world tasks. Thus, the best way to support young people's development of data literacy is not through direct instruction in data-related skills but through exposure to them in applied settings. Since the goal of data literacy is to help individuals learn to illuminate real-world phenomena through data, learning should be project-based, problem-driven, and culturally relevant.

One model for inspiration is the recent evolution of youth computer science learning. An increasing number of young people are becoming expert programmers thanks to technologies and initiatives to make computer science more accessible and appealing to youth. One of the leading resources for young programmers is Scratch, a programming language developed by MIT's Lifelong Kindergarten Lab that makes it possible for youth to develop interactive projects that are meaningful to them without first gaining a high level of technical expertise. Scratch lowers the barriers to participation and enables a wide range of personal expression. In addition, a robust online community of youth, parents, and educators has developed around Scratch—youth comment on each other's work and collaborate on group projects, while parents and educators share experiences and resources for using Scratch at home and in the classroom. We are in a moment when there is great potential for supporting youth in acquiring data literacy. As we begin to do that, we can look to Scratch as a model for how to engage a broad range of youth and educators and sustain that engagement through community.

In the following sections, I will describe a range of methods for supporting youth in the development of data literacy, including technologies that enable youth to interface with data, models for out-of-school data literacy programs, and strategies for integrating data literacy into school curricula.

### **2.2.1 Educational Technologies**

Using new technologies such as sensors, mapping tools, and data visualization tools, which employ simple user interfaces and automate complex tasks, novices can now participate in data collection, interpretation, and visualization, activities that were previously limited to experts. Technology can be used inside and outside school learning to expand the range of opportunities for youth to collect, explore, manipulate, and communicate data. Introducing technology in educational settings provides youth with opportunities for independent exploration and discovery and enables deeper levels of interaction. Furthermore, mobile technologies create new possibilities for interacting with data beyond the confines of the classroom and comparing local and global information. These technologies are a critical component of developing data literacy, but they do not make an impact autonomously; they must be embedded within meaningful learning experiences facilitated by educators or mentors. Below, I will describe four types of educational technology that can be used to support data literacy, and I will provide examples of how they can be used with youth.

Youth can use sensors to collect data about their local environment, constructing their own sensors using electronics kits or using applications that take advantage of the motion, environmental, and orientation sensors embedded in mobile devices. The SENSE IT project, a series of curriculum modules undertaken in 2012 in Washington and New Jersey in which students built a network of sensors to assess water quality, illustrates the potential of integrating sensor technology into a high school science curriculum.<sup>23</sup> During the program, students built, calibrated, and tested a set of sensors and circuits to measure

water temperature, conductivity, turbidity, and depth. Students deployed their sensors, wrote simple computer programs to visualize the data they collected over time, and learned about the larger environmental implications of their results. Finally, students learned to create a wireless sensor network composed of many individual sensor nodes in order to automate data collection. The program illustrates how students can use environmental sensors to study a local problem and local data, supporting data literacy through data collection and data analysis, thereby connecting data to larger principles.

In another example, the ability to understand geographic data, historically supported through the use of maps and charts, is now enhanced by interactive technologies. By studying geographic data using mapping software, youth can discover relationships between geography and society and learn to interpret environmental and social-spatial patterns. Furthermore, Geographic Information Systems (GIS) software enables users to visualize data geographically by creating their own maps. Due to the complexity of GIS software, it is often used simply to view and analyze data in introductory settings. Yet the recent development of tools like ARC GIS, Google Earth, and Social Explorer, which provide simplified interfaces for visualizing and exploring geographic data for people who are not GIS experts, have made mapmaking accessible to a broader audience. These tools give a great amount of control in exploring and manipulating complex data through user-friendly interfaces. The Data-Enhanced Investigations for Climate Change Education (DICCE) project, for example, provides youth with a user-friendly tool for interfacing with a database of Earth observation data collected by NASA, which they are able to visualize through graphs or maps in order to study climate change. In one use case, a pair of eighth graders used data about air temperature, cloud cover, and rainfall to create a model for predicting flash floods.<sup>24</sup>

Participatory mapping technologies also enable personalized geographic investigations, in which youth contribute their own data that is represented geographically on a map. For example, a curriculum developed by Sarah Elwood and Katharyne Mitchell

for seventh grade students in Seattle used interactive mapping technologies to enhance students' learning of local history by enabling them to research, compare, and discuss the cultural and political histories of their neighborhoods.<sup>25</sup> Students used a custom mapping platform to add their own annotations, comments, and sketches to local historical sites, drawing from online and archival research, interview data, and personal stories. The activity enabled them to gain a better understanding of historical data by relating it to their personal environment and experiences.

Finally, visualization tools provide interfaces for youth to manipulate data and communicate it in various formats. Software and web tools like Excel, Google Fusion Tables, and Tableau make it easy to create a wide range of visualizations from complex data sets, without requiring any programming skills, in order to compare and communicate data. For instance, in 2011 tenth and eleventh grade students at the iSchool in New York City used the interactive data visualization tool Gapminder to analyze changing world demographic data over the last 200 years, answering research questions they developed, such as "How does violent conflict impact economic development in the Democratic Republic of Congo?" and "How does HIV affect education rates in Botswana and Zambia?"<sup>26</sup> Such technologies enable a level of data analysis that was previously inaccessible without a high level of technical knowledge.

### **2.2.2 Out-of-School Learning Models**

Since data literacy spans a wide range of disciplines, there are many approaches to using data to help youth investigate relevant and pressing issues—and, in the process, to build data literacy. The following models for out-of-school learning exemplify how data-related activities can be framed as a way to investigate social issues, train for a potential career path, or build applications with real-world use.

The first model frames data as a window into social issues. Local data collection is used to build awareness around social issues and to involve youth in problem solving for social justice. Youth Participatory Action Research introduces youth to data because it offers “opportunities to study social problems affecting their lives and then determine actions to rectify these problems.”<sup>27</sup> It is centered on social justice and designed to challenge and transform systems and institutions. For instance, in the “Echoes of Brown” project organized by the Public Science Project, youth from racially integrated urban and suburban high schools surveyed their peers to analyze intergroup relations and the “opportunity gap.” Youth learned how to conduct interviews, design surveys, and organize the qualitative and quantitative data they collected. They learned where to look for evidence of educational injustice and contextualized their work by studying the history of the civil rights movement.<sup>28</sup>

The second model focuses on building skills and experience. Data exploration is used as a way to engage youth in activities that will teach them specific skills and excite and prepare them for data-related careers. The Chicago Summer of Innovation, for instance, was a summer jobs program for Chicago teenagers to train youth to use technology and data in journalism. The 150 youth in the program were divided into groups according to their interests and worked with mentors to produce issue-oriented media to reach, educate, and motivate Chicago youth to make healthy choices. During the program, they used a combination of statistical data and interview data to support narratives and communicate those narratives using media. Thus, youth learned to use data within the context of a professional activity. The teenagers in the program were motivated and engaged because they had a sense that the work they were creating was serious and public.

The final model is a project or design approach to working with data. The aim of project-based data learning is to encourage young people to create products through a hands-on, self-directed design process, using data to help support their design goals.

One example of this model, which I describe in more detail in Chapter 3, is the series of youth hackathons organized by UK-based organization Young Rewired State (YRS) to enable youth with existing programming knowledge to create applications using open data and to provide them with the community and support to incubate and realize their ideas. During the hackathons, YRS provides youth access to a large number of data sets from a range of sectors, and data savvy mentors help teams find and sort through data they might want to work with based on their ideas and interests. Youth have developed applications that range from an app to track food trucks in San Francisco, to an app to plan the best route for a heist or getaway in Bristol, to an algorithm for choosing winning and losing defense and prosecution arguments during trials, to a crowd sourced mapping application to commemorate the homes of Jews who were taken during the Holocaust in Berlin.<sup>29</sup> In this model, youth move beyond the basic applications of data visualization, developing creative applications to address issues and topics of their own choosing.

### **2.2.3 Data Literacy in Schools**

Even though the Common Core Standards,<sup>30</sup> have long recognized the value of understanding data within the context of different subject areas, there are insufficient resources in schools for helping youth to develop data literacy. In the typical school curriculum, data sets are typically either borrowed from contexts far outside students' experiences or invented for educational purposes.<sup>31</sup> In either case, these data sets are already clean and orderly, ripe for prescribed analysis. Students usually do not have the opportunity to see how data is collected in the field. This means they do not experience the complexity of processing and analyzing a data set with real world implications. Furthermore, data exploration is typically compartmentalized to a single methodological focus and is rarely found outside a science or mathematics class, boundaries which are extremely limiting.

Implementing a data literacy curriculum in K-12 education is challenging because it requires updates to classroom technology and significant teacher training, and it transcends boundaries of school subjects.<sup>32</sup> This is especially an issue for poorer schools, which lack resources such as technology and capacity for teacher training. Data literacy models must make compromises in order to fit into the school model. One model for overcoming these challenges in the short term is to integrate meaningful data explorations with learning objectives from a specific subject. For instance, the SENSE IT project and Elwood and Mitchell's participatory mapping project, both described in the previous section, offer examples of how data exploration can be integrated into an environmental science class or in a social studies context. City Digits, a project that I worked on with MIT's Civic Data Design Lab in 2013 and discuss further in Chapter 3, illustrates how data exploration can be integrated into a culturally relevant mathematics class at a public school in Brooklyn, NY. Mathematics is one subject area that lends itself to data exploration. The new Common Core State Standards for Mathematics recognizes the importance of such skills.<sup>33</sup> It emphasizes interpreting data, recognizing trends, making inferences, and justifying conclusions based on data, within an overall approach that promotes applying critical thinking. This project combined local data collection, data analysis, and construction of data-supported arguments to enable students to investigate the social implications of state lotteries. By pairing active real-world data collection with data analysis, City Digits helped students connect abstract concepts with concrete issues relevant to their own lives. During these activities, topical mathematics concepts were used to support data exploration.

Another strategy to support data literacy within the constraints of the school system is to separate learning across multiple classes. One project that illustrates this model is the Thinking with Data project, a series of four two-week modules for seventh-grade social studies, mathematics, science, and English language arts classes undertaken in 2008 by the Kent State Research Center for Educational Technology in collaboration with SRI

International. The issue of water allocation was chosen as a topic to unite data learning across the curriculum. In the social studies module, students analyzed a variety of data in order to learn about issues surrounding the fair allocation and use of water in the Tigris-Euphrates watershed. Next, students learned about proportional reasoning during the mathematics module as preparation for the following modules. In the science module, students learned to use data to identify patterns and relationships, applying these skills to address the science behind water use debates. Finally, in English language arts, students learned to synthesize and communicate their findings about the water crisis.<sup>34</sup> Through this model, students built data literacy skills and gained an understanding of data literacy's cross-disciplinary nature. They had the opportunity to focus on different facets of data literacy in depth, which is not always possible within the context of a single class.

The examples I have mentioned illustrate the potential of integrating data literacy programs into school curricula. Yet these individual examples are not typical, and expanding them is a considerable challenge since public education is resistant to widespread change. Despite the challenges of integrating data literacy programs into schools, I believe that schools play a critical role in preparing youth to succeed in a society heavily impacted by data and should do more to foster data literacy. While out-of-school programs have more freedom in developing data literacy programs, they lack schools' reach, especially in terms of engaging underserved students who might not participate in extracurricular activities. Schools should embrace and value data literacy curricula because of their interdisciplinary and exploratory nature, not in spite of it.

## NOTES

- 1 Raymond Williams, *Keywords: A Vocabulary of Culture and Society*, Rev Sub edition (New York: Oxford University Press, 1985).
- 2 Media scholar David Buckingham points out that literacy is often used to refer to abilities rather than critical thinking: “Literacy’ comes to be used merely as a vague synonym for ‘competence’, or even ‘skill.’” David Buckingham, “Digital Media Literacies: Rethinking Media Education in the Age of the Internet,” *Research in Comparative and International Education* 2, no. 1 (January 1, 2007): 43.
- 3 Jacob Carlson et al., “Determining Data Information Literacy Needs: A Study of Students and Research Faculty,” *Portal: Libraries & the Academy* 11, no. 2 (April 2011): 25–26.
- 4 Javier Calzada Prado and Miguel Ángel Marzal, “Incorporating Data Literacy into Information Literacy Programs: Core Competencies and Contents,” *Libri: International Journal of Libraries & Information Services* 63, no. 2 (June 2013): 128–129, doi:10.1515/libri-2013-0010.
- 5 <http://schoolofdata.org>.
- 6 Sonia Livingstone, Elizabeth Van Couvering, and Nancy Thumin, “Converging Traditions of Research on Media and Information Literacies: Disciplinary, Critical, and Methodological Issues,” in *Handbook of Research on New Literacies* (New York: Lawrence Erlbaum Associates/Taylor & Francis Group, 2008), 107.
- 7 Ibid., 108.
- 8 Lynn Arthur Steen, *Mathematics and Democracy: The Case for Quantitative Literacy* (Woodrow Wilson Natl Foundation, 2001).
- 9 Lynn Arthur Steen, 1941- Steen, “Numeracy: The New Literacy for a Data-Drenched Society,” *Educational Leadership* 57, no. 2 (October 1999): 8–13.
- 10 Rachel Schutt, “Taking a Chance in the Classroom: Embracing the Ambiguity and Potential of Data Science,” *CHANCE* 26, no. 4 (2013): 46–51.
- 11 Jeannette M. Wing, “Computational Thinking and Thinking about Computing,” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 366, no. 1881 (October 28, 2008): 3717–3725, doi:10.1098/rsta.2008.0118.
- 12 Valerie Barr and Chris Stephenson, “Bringing Computational Thinking to K-12: What Is Involved and What Is the Role of the Computer Science Education Community?,” *ACM Inroads* 2, no. 1 (February 2011): 48–54, doi:10.1145/1929887.1929905.
- 13 Livingstone, Van Couvering, and Thumin, “Converging Traditions of Research on Media and Information Literacies: Disciplinary, Critical, and Methodological Issues,” 105.

- 14 Brian V. Street, *Social Literacies: Critical Approaches to Literacy in Development, Ethnography and Education* (Routledge, 1995).
- 15 A 2013 Pew survey indicates that eighty percent of American adults are able to access the Internet at home through a broadband connection or smartphone. Aaron Smith, "Home Broadband 2013," *Pew Research Center's Internet & American Life Project*, August 26, 2013, <http://www.pewinternet.org/2013/08/26/home-broadband-2013/>.
- 16 "Big Data: The next Frontier for Innovation, Competition, and Productivity | McKinsey & Company," accessed February 16, 2014, [http://www.mckinsey.com/insights/business\\_technology/big\\_data\\_the\\_next\\_frontier\\_for\\_innovation](http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation).
- 17 "Gapminder," accessed April 30, 2014, <https://sites.google.com/a/nycischool.org/gapminder/about-gapminder-at-the-ischool>.
- 18 Seymour Papert, *The Children's Machine: Rethinking School In The Age Of The Computer*, Reprint edition (Basic Books, 1994).
- 19 Allan C. Ornstein and Francis P. Hunkins, *Curriculum: Foundations, Principles, and Issues (6th Edition) (The Allyn & Bacon Educational Leadership)*, 6 edition (Pearson, 2012).
- 20 Urs Gasser et al., *Mapping Approaches to News Literacy Curriculum Development: A Navigation Aid*, SSRN Scholarly Paper (Rochester, NY: Social Science Research Network, November 14, 2013), <http://papers.ssrn.com/abstract=2354500>.
- 21 John Dewey, *The Child and the Curriculum: And The School and Society* (Chicago: University of Chicago Press, 1956).
- 22 Urs Gasser et al., *Mapping Approaches to News Literacy Curriculum Development: A Navigation Aid*, SSRN Scholarly Paper (Rochester, NY: Social Science Research Network, November 14, 2013), 4, <http://papers.ssrn.com/abstract=2354500>.
- 23 Liesl Hotaling et al., "SENSE IT: Teaching STEM Principles to Middle and High School Students through the Design, Construction and Deployment of Water Quality Sensors," *Advances in Engineering Education* 3, no. 2 (Summer 2012): 1–34.
- 24 Daniel Zalles, "Young Youth Explore Geospatial Data for Citizenship Project: A Case Study" accessed February 19, 2014, [http://sriinternational.com/sites/default/files/publications/01\\_youth\\_explore\\_article.pdf](http://sriinternational.com/sites/default/files/publications/01_youth_explore_article.pdf).
- 25 Mitchell, Katharyne, and Sarah Elwood. "Engaging students through mapping local history." *Journal of Geography* 111.4 (2012): 148–157.
- 26 "Gapminder," accessed April 30, 2014, <https://sites.google.com/a/nycischool.org/gapminder/about-gapminder-at-the-ischool>.
- 27 Julio Cammarota and Michelle Fine, *Revolutionizing Education: Youth Participatory Action Research in Motion*, 1 edition (Routledge, 2008).

- 28 María Elena Torre and Michelle Fine, “A Wrinkle in Time: Tracing a Legacy of Public Science through Community Self-Surveys and Participatory Action Research,” *Journal of Social Issues* 67, no. 1 (March 2011): 106–121.
- 29 <http://hacks.youngrewiredstate.org/>
- 30 The Common Core State Standards is an education initiative in the United States that aims to make education standards consistent across states by setting objectives for what students in grades K-12 should learn in English language arts and mathematics in each grade. The Common Core State Standards have been adopted in 46 of 50 states, in part due to the federal financial incentives established by the Race to the Top program. “Understanding the Common Core Standards,” *Education Digest* 79, no. 8 (April 2014): 16.
- 31 Vivian Lim et al., “Local Lotto: Mathematics and Mobile Technology to Study the Lottery,” in *Cases on Technology Integration in Mathematics Education*, ed. D. Polly (Hershey, PA: IGI Global, 2014).
- 32 Mark Van ’t Hooft et al., “A Cross-Curricular Approach to the Development of Data Literacy in the Middle Grades: The Thinking With Data Project,” *Middle Grades Research Journal* 7, no. 3 (Fall 2012): 19–33.
- 33 National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common Core State Standards for Mathematics*. Washington, DC: Authors.
- 34 Van ’t Hooft.



## CHAPTER 3

# Two Initiatives for Building Socially Embedded Data Literacy

In Chapter 2, I described a range of methods for supporting data literacy. In this chapter, I take a closer look at two data literacy programs for youth. These case studies, I argue, illustrate the enormous potential for programs to engage youth in data exploration as a way to build data literacy; they also illustrate the challenges and limitations of implementing socially embedded data literacy programs in out-of-school and in-school learning environments. I argue that in the long-term, these projects have the potential to scaffold participants' civic participation, support participants' motivation to pursue data-related careers, and broaden participants' interest in engaging with data. The first case study, Young Rewired State's Festival of Code, is a weeklong hackathon for young programmers that introduces them to open data and supports them in building websites, mobile apps, and algorithms using data. The second case study, City Digits, is a high school math class in which youth investigate local social justice issues by collecting and analyzing data. I worked to develop the City Digits project as a Research Assistant for MIT's Civic Data Design Lab.

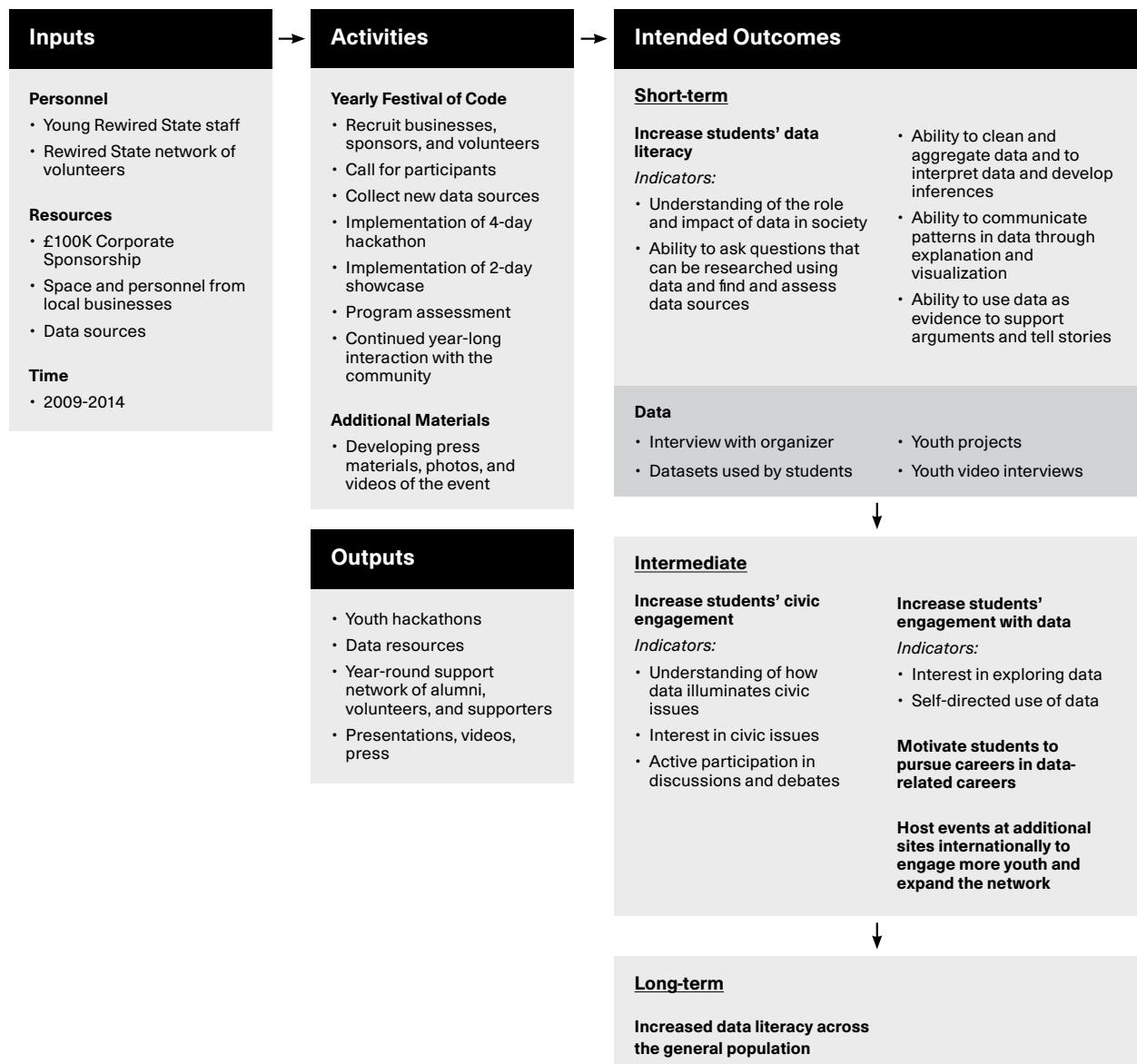
These programs were selected because both exemplify an approach to data literacy that is embedded in students' own social and cultural context. Further, both programs

encourage youth to work with both quantitative and qualitative data and emphasize civic engagement—something I believe we should strive to support through data literacy. Yet these projects were executed using very different methods and in very different contexts. The Festival of Code is an out-of-school program in the UK that introduces youth to data through programming, targets highly self-motivated youth who have taught themselves to code, and engages over a thousand youth every year. City Digits, on the other hand, is an in-school program in New York City that introduces youth to data through local data collection and statistical data analysis; the program, newly piloted with 100 youth in 2013, is aimed at underserved youth, many of whom are struggling in school.

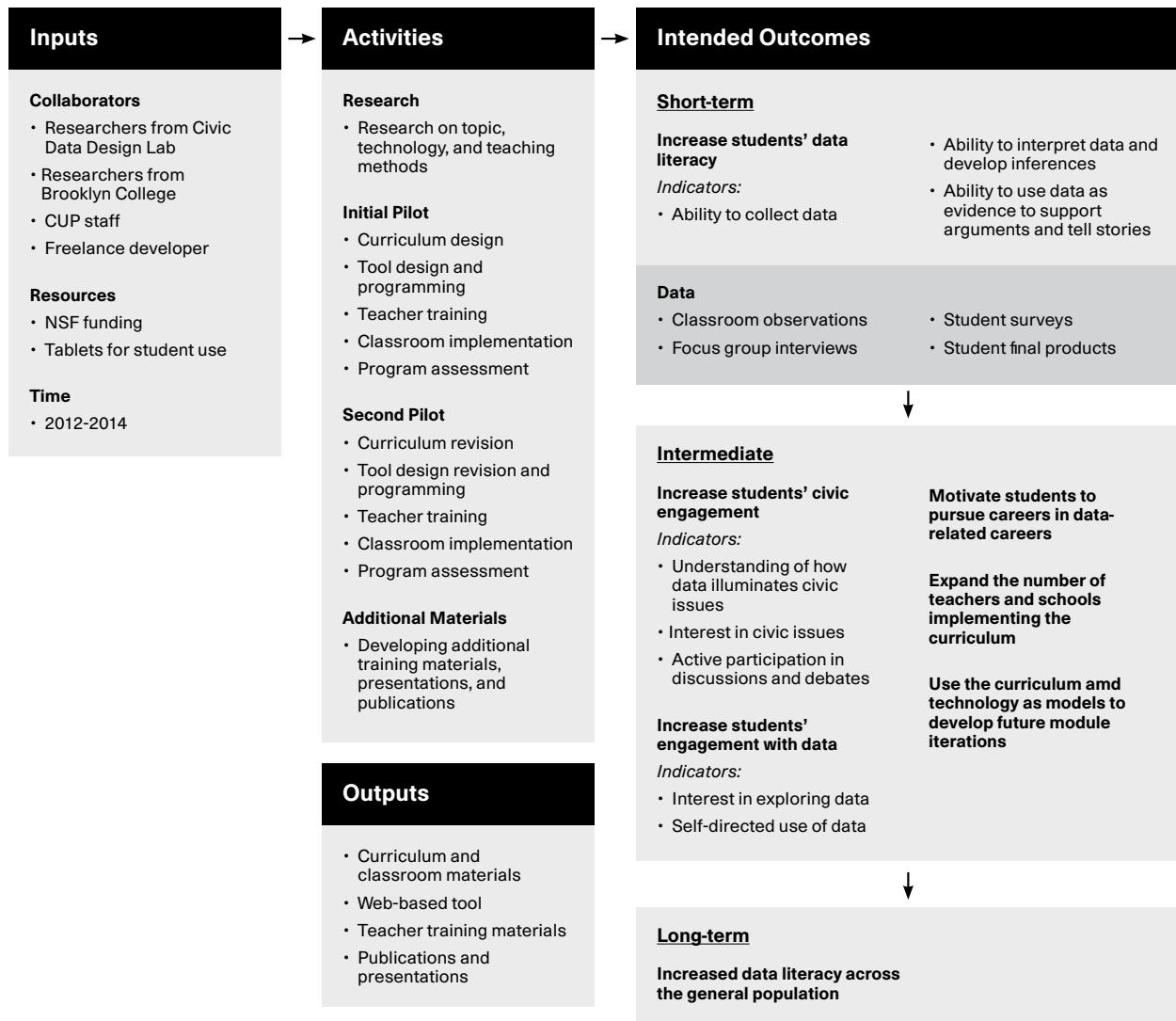
To aid in the analysis of these two programs, I will look for a loose set of short-term indicators, derived from my definition of the data literacy competencies as described in Chapter 2. These indicators will be adapted to each case according to its objectives, as illustrated in Fig. 3.1 and Fig. 3.2. I will look at the ways in which these programs impacted students' understanding of and engagement with data, as demonstrated through a number of qualitative indicators. In addition, I will present anecdotal evidence that suggests these programs have the potential to support intermediate-term objectives of increasing participants' active citizenship, career-related skills, and continued engagement with data. Since Young Rewired State's Festival of Code is in its sixth year, and thousands of youth have attended its events, there is more evidence of its intermediate-term impact—I examine how participants have developed their skills from year to year and what they have gone on to achieve after the event. The Festival of Code serves as an example of the large impact data literacy initiatives can have given the right conditions and resources. City Digits, on the other hand, is barely one year old, and the two pilots were too small and experimental to result in a reliable measure of intermediate-term impact. However, the City Digits case study serves as a useful illustration of the challenges that researchers and educators must face incorporating a data literacy curriculum into a public school environment and working

with less academically motivated youth, which is critical for spreading data literacy to underrepresented communities. It also illustrates the potential impact of this type of project; in the classroom, I witnessed firsthand the short-term transformation that took place as the program deeply engaged a tough to reach group of students.

**Fig. 3.1** Logic Model of Young Rewired State's Festival of Code



**Fig. 3.2** Logic Model of City Digits: Local Lotto



## 3.1 Young Rewired State's Festival of Code

Young Rewired State (YRS), the philanthropic branch of Rewired State, is a UK organization that brings together youth who have taught themselves to program for day and week-long hackathons; at these hackathons, youth learn about open data and work in teams to develop websites, mobile apps, and algorithms using data. The largest event, the

Festival of Code, is a yearly weeklong event hosted at numerous local venues—mostly tech companies, schools, and start-ups—in regions across the UK. Youth attend the venue closest to their home, so there is always a diversity of levels and ages at each location. Youth code from Monday through Friday, and on Saturday and Sunday participants from all regions gather in a single location for a showcase of the work produced over the course of the week.

Many young people attend year after year, so the weeklong hackathon serves as both a means to introduce new young people to data and also a way to bring the existing Young Rewired State community together. For many of the young people, the greatest appeal of the annual event is its potential for establishing friendships and connecting with others who share a passion for code.

This case study draws from interviews with Young Rewired State founder Emma Mulqueeny, documentation of participants' projects, participants' blog posts and tweets, and YRS's video documentation of participant interviews and presentations in order to argue that YRS supports youth in developing data literacy competencies through an open-ended framework that encourages creativity and expression as well as building possibilities for future engagement with data.

### **3.1.1 Background**

Emma Mulqueeny founded Young Rewired State (YRS) in 2009 because she felt it was important to include youth in the dialogue around open data. At the time, the concept of open data was just coming into the public consciousness after a public expenses scandal had shaken citizens' confidence in government accountability.<sup>1</sup> Mulqueeny began hosting adult hack days to help ministers recognize the value of opening government data and to engage programming communities in open data innovation. The events were a success—ministers began to understand the value of open data, developers were excited to hack with government data, and a few months later the government began

releasing data through an open data portal, data.gov.uk. But the hack days were limited to a homogenous audience: “middle-aged grumpy developers.” Mulqueeny wanted to begin including young people. She felt it was important for youth to learn to consume and transform data, since over their lifetime the government would increasingly release information in data form.<sup>2</sup>

Mulqueeny hosted the first youth hack day at Google’s London office in 2009 for 50 young people between the ages of 15 and 18. It was a challenge for Mulqueeny to find 50 participants interested in programming—only three responded to her original call for entries. At the time, most schools she talked to did not have coding classes or clubs. One by one she found isolated kids who had taught themselves to code, and she invited them to participate in the event. Mulqueeny quickly realized that to successfully introduce youth to data, she would first have to establish and cultivate a network of young people who loved to code.

Mulqueeny modeled the youth hack day after the adult hack days she runs—youth were given access to a large number of datasets and were given open ended challenges to guide their work. Mentors, who were open data developers from the Rewired State network, worked with teams of young people to help them learn to work with data. Repurposing the adult hack day model worked well with this audience of young self-taught programmers. By the end of the weekend, “lifetime friendships were born and [the young coders had] fallen in love with open data.”

### **3.1.2 Program Overview**

Since 2009, YRS has grown and refined its model for running its youth hackathons, focusing in particular on its largest annual event, the weeklong Festival of Code. In 2013, over 1,000 youth participated in the Festival of Code, spread across 40 centers. Six-hundred of those youth gathered for the final showcase, hosted at the Custard Factory in

Birmingham. Four weeks before the Festival begins, YRS publishes a repository of datasets for the youth to explore so they can begin to develop ideas before the event begins. This resource includes links to public sector data such as data.gov.uk as well as local data sets and commercial data sets provided by corporate partners.

At the beginning of the week, the participants at each center introduce themselves and talk about their programming experience. Youth who have already explored the data divide into groups, and a mentor helps them talk through their ideas and search for additional datasets. The role of the mentor is to help youth focus their ideas and objectives, to teach them how to navigate data, and to help them differentiate between incomplete data and complete data that will be interesting to interrogate. For youth who are new to the event or have not yet browsed the data, mentors begin by helping to identify an area of opportunity or a problem they would like to address—often something that frustrates them or affects their families. Some of the first-time participants, who have trouble deciding what they want to do or are reluctant to engage, join existing groups or float around the room. By the end of the first day, all the groups have a problem that they want to solve and an idea of the data they want to use.

Joel Murphy, 19, who participated in the Festival of Code in 2011 and 2012 and returned as a mentor in 2013, describes in a blog post how the 18 youth at the Box UK center in Cardiff split into four teams according to their level of experience:

1. The people who had never programmed before worked together to learn HTML and CSS from scratch, before designing a web page for the Welsh Government. This project was more of an educational experience than anything else, but it was an excellent opportunity for the team to dive into the world of code.
2. The younger people who had programmed before decided to create a Team Fortress stat comparison tool, written with a Python back-end and HTML front-end, that tells you which player is best by doing a bunch of calculations based on their Steam gaming history.
3. The slightly older people who had programmed before built an app for managing plant growth, using a variety of programming languages including a Ruby back-end, Ajax polling, C, and a bit of Python to convert data streams from

an Arduino to a bridged Raspberry Pi. 4. The eldest people taking part in the event worked on an app for improving ambulance response times. We decided to put a 10 year-old PHP coder in the team too, as he probably had the best coding knowledge out of anyone in the team. The app used 2.6 million rows of historical crash data and an in-house algorithm to calculate the risk of road accidents occurring at specified locations across the UK.<sup>3</sup>

At the Pearson office in London, on the other hand, the 9 youth ranging in age from 11 to 17 decided that they did not want to split into small teams and instead chose to work as a single group. After brainstorming around the skills and areas of interest in the room, the team came up with the idea of creating a web app comparing the “best of” and “worst of” locations across the UK in terms of local infrastructure, examining variables such as education and land value.<sup>4</sup>

When students return on day two, they have been processing and brainstorming overnight and come back with a plan. Mulqueeny says, “the energy has changed overnight.” Teams become deeply engaged in their work. During the 2013 Festival of Code, one mentor, @One\_KX, tweeted: “#YRS2013 - help! how do you make coders stop coding and take a break over lunch? They won’t stop typing! #superkeen.” Youth participant @JamesBrimer tweeted, “Morning of the second day and we’ve already got 5000+ lines of code :D #yrs2013.”

Over the course of the week, student teams develop their ideas, dig deeper into the data, and build their websites, mobile apps, and algorithms. There are no fixed methods or lessons shaping this process; mentors at each site draw from their own expertise to help guide youth through the development process. Murphy says he introduced the beginners at Box UK to HTML and CSS in just a few hours to get them started, helping them build a template by copying and pasting code from different sources, while advanced groups worked more autonomously.<sup>5</sup> A defining characteristic of the YRS learning experience is that youth engage in peer-to-peer learning with their teammates and within the larger YRS community via Twitter and IRC. The back channels are an integral part of the event’s

dynamic as participants chat and ask for and offer advice—on Twitter alone, there were over 11,000 Tweets from 1551 contributors over the course of the week.<sup>6</sup> For instance, on day two, @PentaSquirkle tweeted “#yrs2013 I NEED HELP WITH XCODE STUFFINGNESS PLEASE CAN SOMEONE HELP ME?” Three minutes later, @EdwardLJParkin responded, “Of course! Add my Skype...and I’ll help.” In many cases, youth have more expertise in a particular area than their mentors. Murphy describes, for example, how Team Growify “put together a neat little electronic circuit that measured soil temperature, humidity, and moisture, as well as the current light level. I personally learnt a lot working with Team Growify. For the first few hours I wasn’t able to help them with their project, as I only had a basic understanding of electronics.”<sup>7</sup> Across the centers, mentors’ role becomes less central over the course of the week as the youth become wrapped up in development.

At the end of the week, young people from all the centers gather together in one location for the weekend to showcase the products they have built (Fig. 3.3). These events are energetic and chaotic. @youngrewired’s tweet from the penultimate day of the showcase captures the atmosphere of the event: “#YRS2013 whoever is putting stickers all over the custard factory, please stop. We are being fined real money now.” The wireless Internet often fails because there are hundreds of youth trying to download large datasets simultaneously, and many teams stay awake all night playing with data and perfecting their apps and presentations. Participant @mellis1995 tweeted, “1. Write code. 2. Realise it doesn’t work. 3. Cry in the corner. 4. Repeat until you and your code is broken forever. #YRS2013.”

During the showcase, youth have the opportunity to present their work and see what others have been doing, and awards are given for the best work (Fig. 3.4). There is work from a broad range of participants shown at these events, ranging from beginners who are just putting data on a map, to experts who are creating sophisticated apps. According to Mulqueeny, the beginners feel comfortable showing their work because they see that

there are others at their level; but they also see the more advanced work, and it serves as inspiration for what they could achieve in the future.

**Fig. 3.3** Participants gather at the Custard Factory



© Young Rewired State

**Fig. 3.4** A participant presents his project at a final showcase event



© Young Rewired State

### **3.1.3 Outcomes**

#### **Short-Term Outcomes: Data Literacy Competencies**

Since my main data points in analyzing YRS were an organizer interview, participant presentations and Tweets, and journalistic interviews with participants conducted by YRS—not in-depth research interviews or pre- and post-assessments—I cannot offer a precise picture of what participants learned from the experience versus what knowledge that they came in with. However, participants' excitement about data and the sophistication of the projects they created using data suggest that YRS was effective in supporting data literacy.

In the Young Rewired State model, learning is personalized to the objectives of individual groups, and thus not all students acquire exactly the same skills. Small group interactions with an expert ensure that youth are supported in developing the skills they need when they need them. At the same time, adults play the role of mentor, not teacher, so they help direct self-guided learning. This model works well for these youth, most of whom are self-taught programmers. In addition, peer learning enables youth to learn from others who are on a similar or slightly more advanced level, empowers youth to act as experts and to learn by teaching, and helps youth to practice collaborating with a team.

A series of twenty-nine YouTube videos produced by YRS, capturing journalists and YRS staff interviewing numerous YRS participants, provide evidence that participants gain an understanding of the concept of open data and an appreciation of the potential for innovating with data. In one video, Ben Nunney from Twilio, one of the companies YRS partnered with, conducts journalistic interviews with youth at the Ipswich, Bletchley, and Cambridge centers to discuss their excitement about data. One teenager explained that, “there’s so much open data out there, especially from the government, and the things you can do with it, the ways you can interpret it, extrapolate more information from it, it has so many possibilities.” Another argued that open government data is important because “it

can sort of like help with innovation instead of keeping everything closed and allows people to actually see what's happening in their government." A third teen asserted that these participants "will be, like, the people who are running the country in the future, and most of it will be done with the newest technology, which you need to know code for if you want to develop it. So it's kind of the future."<sup>8</sup>

On the other hand, Mulqueeny says that when youth begin to understand the beauty and potential of data, their excitement is often accompanied by frustration upon discovering that datasets are not always neat and refined. There is a discovery curve: on the second and third days, youth really begin digging into the datasets and find data that interests them, but they often discover that the data they want to use is messy or incomplete. Participants' tweets about data range from excitement—@nickiehills: "Wow, APIs for UK police data access already. Impressed - documentation is here: <http://t.co/RZUWcn4lsS> #YRS2013 via @ayymanduh"—to frustration—@naythdunn: "the Department of Education website is so confusing! -\_- #YRS2013"; @barnaby\_taylor: "I already hate the google maps API #yrs2013 #caerus." These experiences illustrate the process of learning to find, assess, and prepare datasets for use.

Ultimately, the strongest evidence of the data literacy competencies that participants developed is the range of projects they created using data. All of the 125 projects demonstrate multiple data competencies. For instance, Growify, a web app to manage plants and optimize growing conditions, illustrates the team's ability to collect data via sensors using Arduino and Rasberry Pi; Mighty Flight, a map tool that visualizes global flight connections, illustrates the team's ability to visualize geographic data; Mortune, a music recommendation service, demonstrates the team's ability to algorithmically analyze and group data; and re:ACT, a site that allows users to discover, track, and debate bills and legislation going through Parliament, offers a platform for others to understand and react to data.

### **Intermediate-Term Outcomes: Active Citizenship**

During the Festival of Code, the YRS youth work with a wide range of commercial and public data applied to a variety of domains, so civic engagement is not an explicit part of the program design. However, many participant groups chose to use the hackathon as an opportunity to take action on pressing social issues. For instance, many teams chose to create apps that benefit their own communities: one platform maps safe routes to school; another helps local councils and residents fill vacant properties. Others chose to focus on current topics they care about: one app enables users to shoot “evil” Ministers of Parliament with a Nerf gun; another helps users track their daily carbon footprint.

The possibility for wider engagement is illustrated by one group at the 2013 YRS youth hackathon in Berlin, composed of three 16 and 17 year olds from different parts of Germany who met at the event and developed an app called PlateCollect. These youth worked with a dataset that contained the locations of 5000 “Stolperstein,” or tripping stones, which were set into the streets of Berlin to commemorate the former residences of Jews who were deported to concentration camps during World War Two. During the hackathon, the students first created an iPhone application showing the location of all the plaques and links to information about the victims. Through historical research, they realized that there must have been many unrecognized Jews taken from Berlin who were not commemorated by the plaques, so they built a platform to crowdsource the missing information.

By developing an app that was not simply about data consumption but also enabled data collection, these youth created the potential to actively engage others in commemorating the victims. The idea also resonated across cultures—Mulqueeny said that when the group presented their project at the European Commission’s ICT 2013 Conference, they were approached by numerous delegates who could see the potential for similar projects in their own countries, including a woman from Kosovo who wanted to commemorate people who had gone missing during the Kosovo War. Mulqueeny wants to help realize this vision by

connecting the youth from Berlin remotely with youth in Kosovo and other countries.

Through the Festival of Code and other YRS events, youth are prepared and inspired to engage in future opportunities to develop civic technology. For example, in 2012 a team of youth who had participated in the Festival of Code for three years attended a Rewired State charity hack day for Refugees United. The organization sought a better way to reunite Somali refugee families separated in camps across Kenya. Refugees could not be matched by name, because many had common names, their names sometimes changed, and they were reluctant to give identifying information. This challenge was compounded by a lack of access to technology. The team created a feature phone app that makes it possible to match refugees with their families in full anonymity. Refugees share stories that only people from their village or family would know. The algorithm collects these stories into a database and parses them to match people who tell the same stories.

Through these data innovations, youth recognize their own agency—their capability not only to engage with civic issues but also to take action and make an impact. The previous examples demonstrate the potential for youth work to be more than an exercise in civic engagement: these young people created serious tools for civic action. Since YRS has very limited data on youth before and after they participated in YRS, open questions remain as to whether YRS was a catalyst for these behaviors or simply gave youth an opportunity to express existing interests, and whether participating in the program has impacted their engagement in other forms of civic behavior. A subject for future research will be to discover whether participants are becoming more engaged with changing their local communities or influencing their local government.

### **Intermediate Term Outcomes: Continued Engagement with Data and Career Preparation**

Young Rewired State has begun to see youth who, after participating in the Festival of Code, go on to independently develop their skills and return to the event for multiple years.

One of the reasons for this is the sense of community and competition the event inspires. According to the event organizers, when participants attend the final showcase and see the winning teams' advanced work, they become motivated to deepen their skills before the next year's event. Zak, who is 8 years old, was the youngest member of the Contag.io team, which developed an app to model the spread of a virus in a school environment based on data by researchers at Stanford (Fig. 3.5). The app won "Best Example of Code" at the 2013 Festival of Code. When Zak attended the 2012 Festival of Code at the age of 7, he was just beginning to learn to code and was disappointed that he did not win a prize. Seeing the winning projects at the showcase sparked his competitive spirit. When he returned in 2013 he had developed his programming skills and strategized to team up with the most advanced programmers at his center in Manchester. Zak explained that he loves to code because the computer is "kind of like a new place, where you can make whatever you want."<sup>9</sup>

**Fig. 3.5** Zak and his team presenting the Contag.io project



© Young Rewired State

Through YRS youth also become members of a tight-knit community of interest that extends beyond the annual event. Through continued interaction with the network,

youth receive inspiration, encouragement, and support to continue their work and expand their knowledge. This support, however, is generally self-sustained by participants—YRS’s mission is to host hackathons and make connections, not to formalize or direct this community. As a result, YRS works best for youth who are able to acquire skills independently; those who need additional support must seek it elsewhere.

There are also examples of participants whose experiences at YRS have directly led to career opportunities. Mulqueeny says that some YRS alumni, who met at the event and worked together in teams, have gone on to start businesses together. Alex, an 18-year-old who participated in the Festival of Code for three years in a row, took a year off from college to work for Microsoft, an opportunity he was offered by a connection he made at the Festival of Code. He said the skills he developed at the event were directly applicable to this work: “The first year I could code, but I didn’t know very much about handling data, so I learned how to do that. And I learned how to use the actual websites the government uses to host it. And the second year I learned much more about teamwork and working in a team to get a project done on time.”<sup>10</sup> These career connections are also self-generated; the role the YRS hackathons play is to provide participants with an opportunity to interact with a resource-rich community and build relevant job skills.

### **3.1.4 Conclusion**

The outcomes of Young Rewired State’s Festival of Code demonstrate this model’s ability to aid young programmers in building data literacy, engaging with civic issues, and sustaining deep engagement with data through interaction with a community of peers. When youth return to the event year after year, in addition to receiving continuous community support, they are motivated to independently build new skills and give back to the community by encouraging new participants. Furthermore, YRS aims to be inclusive by eliminating barriers to attendance. It offers free admittance, finds local business partners to host

hackathon venues in any regions where there is demand, and covers the cost of hotels and food for participants who travel to the final showcase. Its scale will likely grow in future years since the UK recently mandated that programming become part of the elementary school curriculum starting in 2014.

The flaw in the YRS model is that it does not engage youth who have not had the opportunity to learn to code, who are not self-motivated learners, or who do not self-identify as having enough coding expertise to attend. Further, the dramatic gender gap that we see in computing fields is mirrored within the group of YRS attendees. Indeed, one of YRS's greatest challenges has been to recruit girls. Although Mulqueeny's efforts on this front have succeeded in increasing the number of female participants in recent years, from 3% to 18%, boys still outnumber girls at YRS events, because more boys code than girls. Although YRS does not collect data about participants' race or family income, we know that lower-class and minority communities are also traditionally underrepresented in computing fields because they are less likely to have access to the resources and support necessary to develop coding skills. It is reasonable to expect that this gap would also be mirrored in the YRS community. In looking to this case as a model for future data literacy endeavors, it will be important to think more about how to include youth who are excluded because of deep social barriers that have prevented them from developing coding skills, and to ask how to engage beginners and communities who might not be independently motivated to participate. This is a challenge I will explore further in Chapter 4, "Data Literacy Challenges and Design Principles."

## 3.2 City Digits: Local Lotto

The City Digits team is developing two curricular modules, each around a local social justice issue, and building accompanying geospatial technologies that enable high school

students to explore their local urban landscape, collect field data, and identify and analyze patterns. The first module, Local Lotto, focuses on the theme of the social impact of the lottery and was piloted at a public high school in Brooklyn, NY. The team is currently developing a second module on the theme of pawn shops and fringe banking in New York City. The modules are designed to be implemented in New York City high schools in low-income areas, where mathematics scores are persistently low, in an attempt to ground mathematics and data concepts in real-world issues, to make them more engaging and intuitive, and to encourage students to take an active role in exploring issues affecting their communities.

As a Research Assistant for MIT's Civic Data Design Lab, I worked with a team to develop the City Digits: Local Lotto curriculum, to design technologies for integration into the classroom, and to implement the curriculum and technology in the classroom during two rounds of piloting. This case study draws on observations from the design process, notes taken during class observations, pre- and post-pilot interviews with a student focus group, student surveys about their learning experiences, and final products that the students created during the class.<sup>11</sup> Ultimately, I will use the case study to argue that City Digits serves as a model for overcoming the challenges of supporting data literacy within the constraints of the public school system and engaging underserved youth in data exploration.

### **3.2.1 Background**

City Digits is a collaboration of interdisciplinary partners from three organizations: mathematics education researchers from the City University of New York's Brooklyn College; informal learning environment designers from the Center for Urban Pedagogy (CUP), a not-for-profit education and advocacy organization; and civic technology researchers from MIT's Civic Data Design Lab.<sup>12</sup> Each of the partner organizations approached the project with a distinct but complementary vision shaped by its own

experience working with youth and technology. The Civic Data Design Lab sought to integrate technology into the curriculum to create new opportunities for bringing together in-classroom and out-of-classroom learning and to lower the barriers to data exploration. CUP aimed to engage youth in exploring a local social justice issue through participatory research. And Brooklyn College wanted to introduce culturally relevant mathematics pedagogy into the classroom in order to engage underperforming students. The team also worked closely with high school teachers from the New York City Department of Education, who gave feedback throughout the process and helped pilot the course in their classes.

The City Digits team developed a data literacy curriculum in schools in order to reach underserved students who would be unlikely to enroll in an extracurricular program. A mathematics curriculum worked well as an environment for data exploration since the Common Core Standards for Mathematics encompass skills such as interpreting data, recognizing trends, making inferences, and justifying conclusions based on data.<sup>13</sup> Furthermore, by embedding data literacy objectives within a math class, we were able to test a new model for teaching mathematics.

There is an urgent need for innovative high school mathematics curricula in New York City, where testing scores consistently lag below the national average and have dropped significantly after tougher Common Core Standards were introduced in 2013. Further, in New York City there is a dramatic achievement gap between blacks, Hispanics, students with disabilities, English Language Learners, and students from low-income families and their peers, and there has been little improvement in closing that gap over the last decade.<sup>14</sup> Although standardized testing results do not necessarily capture the nuances of what and how students are learning, or point to solutions, they do indicate a need to rethink the way mathematics is taught in schools. One of the students who participated in City Digits described the problem this way: In typical math classes “the teacher gives you a paper and

explains it, then you probably finish the paper before she's done talking, and you sit and talk for the rest of class. It's boring."

The coordinators of the City Digits project believe that engaging in data exploration that is culturally relevant and grounded in the local environment will give students a new way to look at mathematical problem solving and will help engage students who do not understand or are not interested in traditional approaches. Active investigations within the community will furthermore help build civic awareness. The project combines the theoretical frameworks of place-based education,<sup>15</sup> which builds on students' everyday out-of-school experiences, and culturally relevant pedagogy,<sup>16</sup> which situates learning within cultural contexts of race, ethnicity, and language.<sup>17</sup> The curriculum is based on the idea that combining quantitative analysis of large data sets with active collection of local qualitative data will enable students to examine an issue through multiple lenses, resulting in deeper understanding and more nuanced analysis. In addition, the curriculum explores mobile technology's potential to enable out-of-classroom data collection and to disrupt traditional approaches to mathematics and data learning.

The first City Digits module, Local Lotto, focuses on the theme of state lotteries. The lottery was selected as a topic that presents opportunities for both teaching data concepts and exploring a locally relevant social justice issue. The curriculum and technology were designed around data obtained from the New York State Lottery Commission and public data from the 2010 Census. Together, this data provides numerical evidence on the socio-economic implications of the lottery at a citywide level. We compared 2010 lottery sales and winning data for New York City retail stores that sell lottery tickets to demographic and household income data. The curriculum is structured around the idea that pairing quantitative data analysis of the impact of the lottery with qualitative research on lottery ticket buyers' motives and beliefs, collected through interviews with neighbors, can help students connect large-scale patterns with real-world phenomena.

### **3.2.2 Program Overview**

City Digits: Local Lotto was piloted twice at a high school in Brooklyn, NY over 14 hour-long class sessions. The first implementation was in April 2013 within an “advisory”<sup>18</sup> class of 15 high school sophomores. The curriculum and technology were revised based on the results of the pilot and a larger pilot was conducted in November 2013 with one teacher and his four 12<sup>th</sup>-grade mathematics classes totaling 95 high school seniors. Our pilot school is in one of the city’s poorest neighborhoods; all of the school’s students qualify for free or discounted school lunches. Forty-seven percent are immigrants with English Language Learner status, and the same percentage have special education status or have been held back in school. Moreover, 79% of the students we worked with had not received a college-ready score on NY’s entry-level algebra exam.<sup>19</sup>

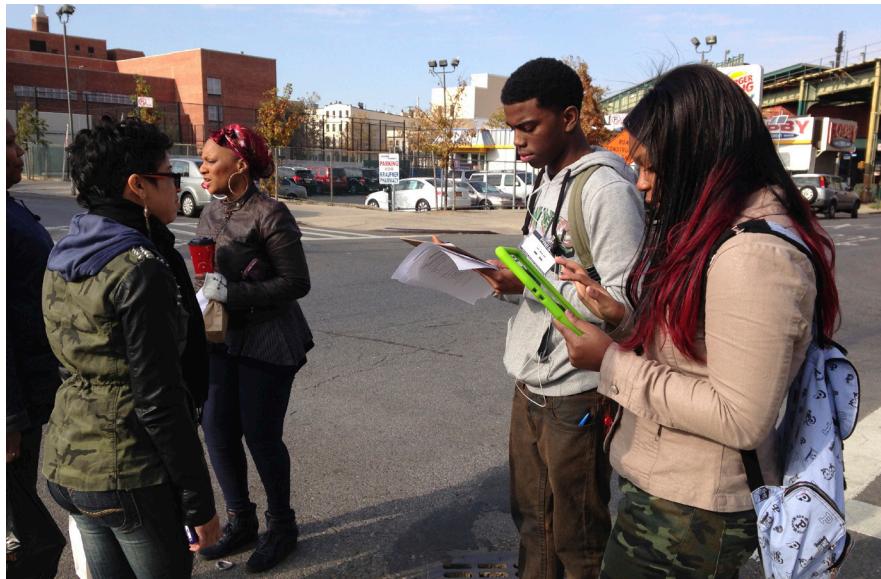
The curriculum is divided into three units, each of which is supported by a custom web application: [citydigits.mit.edu](http://citydigits.mit.edu). First, students collect data about local opinions on the lottery by conducting interviews with pedestrians and lottery ticket sellers at neighborhood stores, using the web application on data-enabled tablets. Second, students analyze citywide lottery data, using an interactive map to facilitate exploration and understanding. Finally, students synthesize qualitative interview data with quantitative map data in order to formulate their own opinions. Finally, using the tool, students create multimedia narratives to teach others about what they have learned.

#### **Local Qualitative Data Collection**

During the first phase of the curriculum, students actively explore the theme of the lottery by interviewing people in their neighborhood about the topic (Fig. 3.6). They become researchers and data collectors rather than passive observers. Small teams of students use the tool on data-enabled tablets while taking various routes around their neighborhood to collect quantitative data about their neighbors’ lottery ticket purchase habits and the

volume of sales at neighborhood stores, as well as qualitative data about their neighbors' opinions about the lottery (Fig. 3.7). Students wanted to find out why and how often people played the lottery and see how local businesses thought the lottery was impacting the neighborhood. Using the tablet allows students to easily navigate, collect data in the field, and instantly publish geo-located results (Fig. 3.8). Students discovered many stories from people for whom the lottery is central, such as this man, who says he spends \$50 on the lottery every week because he cannot find a job and wants to support his family (Fig. 3.9). The field investigations make the issue of the lottery seem more urgent by rooting lessons in neighbors' real stories, while also helping students to acquire data collection and interviewing skills.

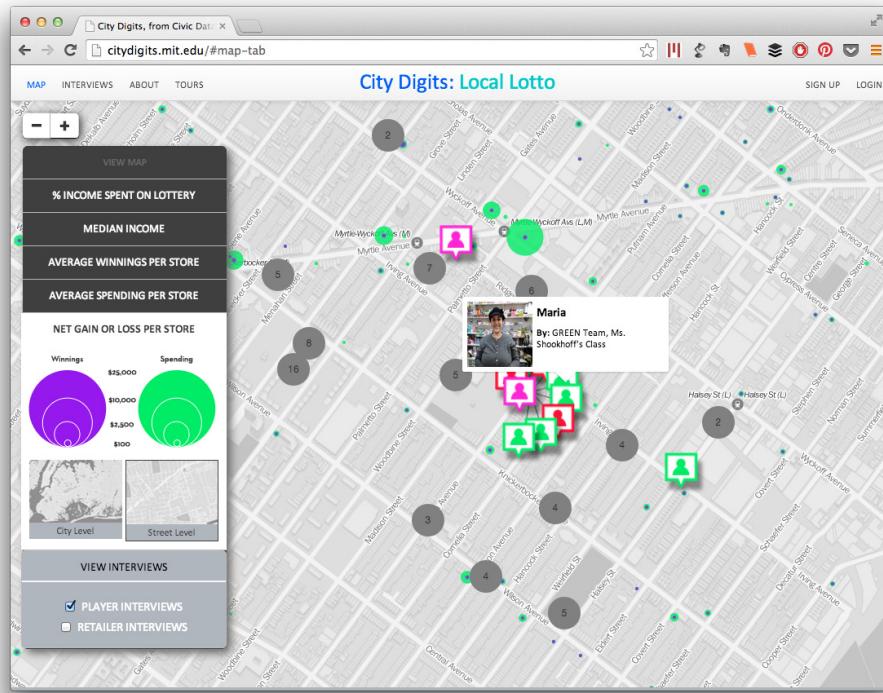
**Fig. 3.6** One group of students conducts an interview with lottery players



**Fig. 3.7** Interview questions for lottery players and retailers

Questions for Lottery Players	Questions for Lottery Retailers
1. Do you ever buy lottery tickets?	1. Do you sell lottery tickets?
2. Why or why not?	2. Why or why not?
3. What is the most you've ever won?	3. About how many customers do you have in an average day?
4. How much do you spend on lottery tickets in an average week?	4. What percentage of your customers buy lottery tickets?
5. What would you do if you won the jackpot?	5. About how many tickets do people usually buy in one visit?
	6. Do you think the lottery is good for this neighborhood? Why or why not?

**Fig. 3.8** Interview points appear on a map that also reveals lottery data about the neighborhood



**Fig. 3.9** This man spends \$50 on the lottery every week



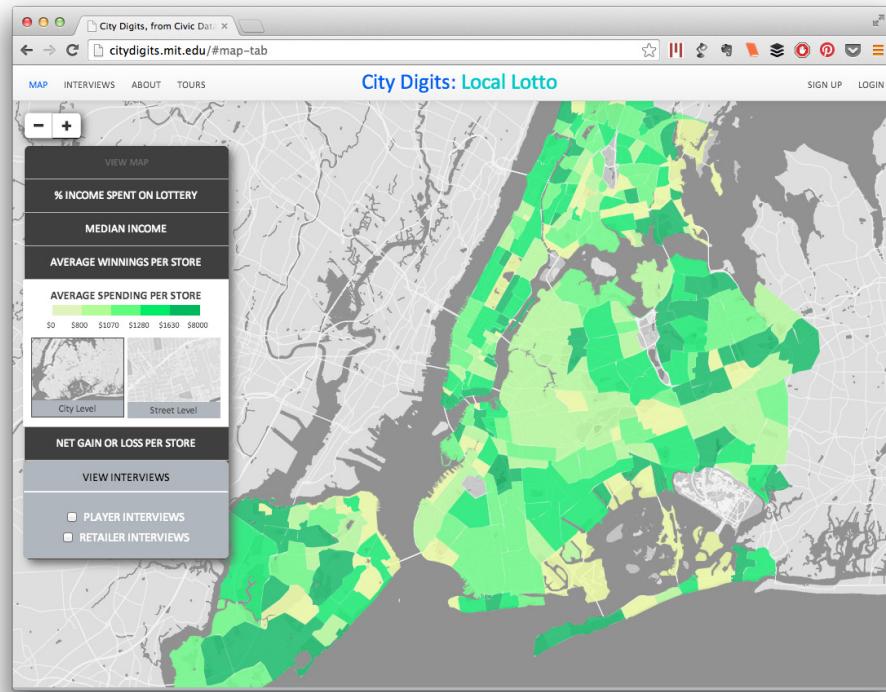
### Citywide Quantitative Data Analysis

In the second phase of the curriculum, students analyze an interactive choropleth map of the city to address questions of the lottery's impact. The map shows lottery spending and winning data both at a citywide level, comparing aggregated data across neighborhoods (Fig. 3.10) and at a highly localized level, comparing individual data from stores that sell lottery tickets (Fig. 3.11). The map also shows lottery spending as a percentage of each neighborhood's median household income (3.12). The visual geographic representation provides an intuitive window into the data, enabling students to identify patterns and analyze the large-scale effects of the lottery as a system by building on geographic knowledge they already have about their city. Students are asked to take into consideration factors that might affect data from different neighborhoods, such as household income, population density, and proportion of residents to visitors. Students are able to use their existing understanding of different neighborhoods to interpret the maps at a more sophisticated level.

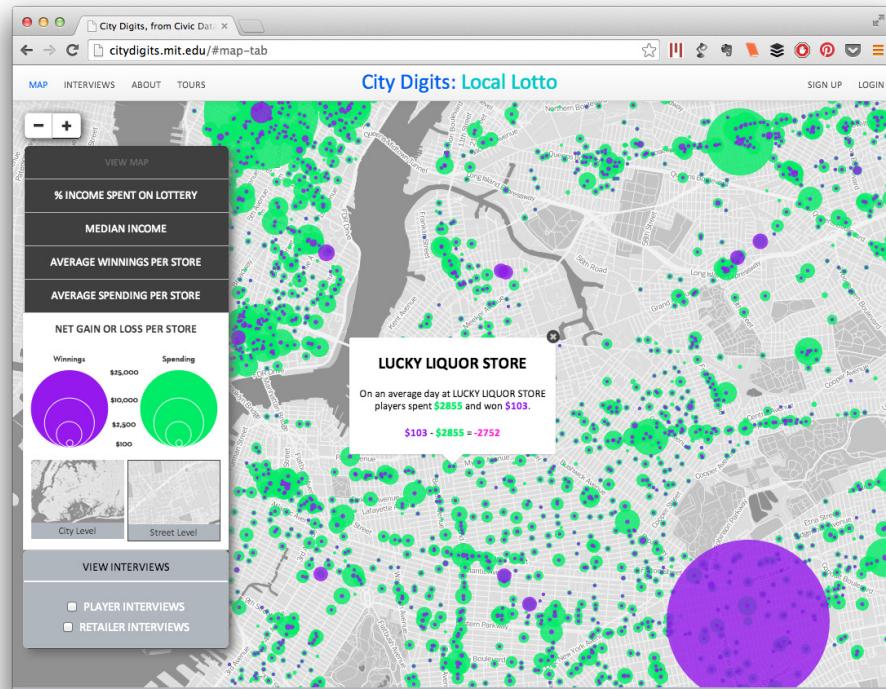
For example, one pattern that we wanted to highlight, and which students picked up quickly, is that lottery players in low-income neighborhoods spend a significantly higher proportion of their income on lottery tickets than players in high-income neighborhoods. This information is recognizable as a citywide pattern and can also be broken down by comparing individual neighborhoods. For instance, in Bushwick the median daily household income is \$100 and adults, on average, spend 3% of their income on lottery tickets. In Park Slope, the median daily household income is \$270, and adults spend only .5% of their income on lottery tickets (Fig. 3.12). By zooming into the street level of the map, students can discover the amount that was spent on and won from the lottery for every store that sells tickets. The data also reveals that, on average, lottery players in all neighborhoods spend much more money on lottery tickets than they win.

After answering a series of exploratory questions in order to get oriented to the map, students formulated their own questions to explore the data. Since the map displays socioeconomic data, much of the student discussion focused on the socioeconomic implications of the lottery, without the teacher having to explicitly prompt students. One topic students were interested in, which was not something we had explored while developing the maps and the curriculum, was looking at the stores and neighborhoods where there had been winning jackpot tickets. Some of the students were interested in knowing more about the winning tickets and went on to research the stories behind the data.

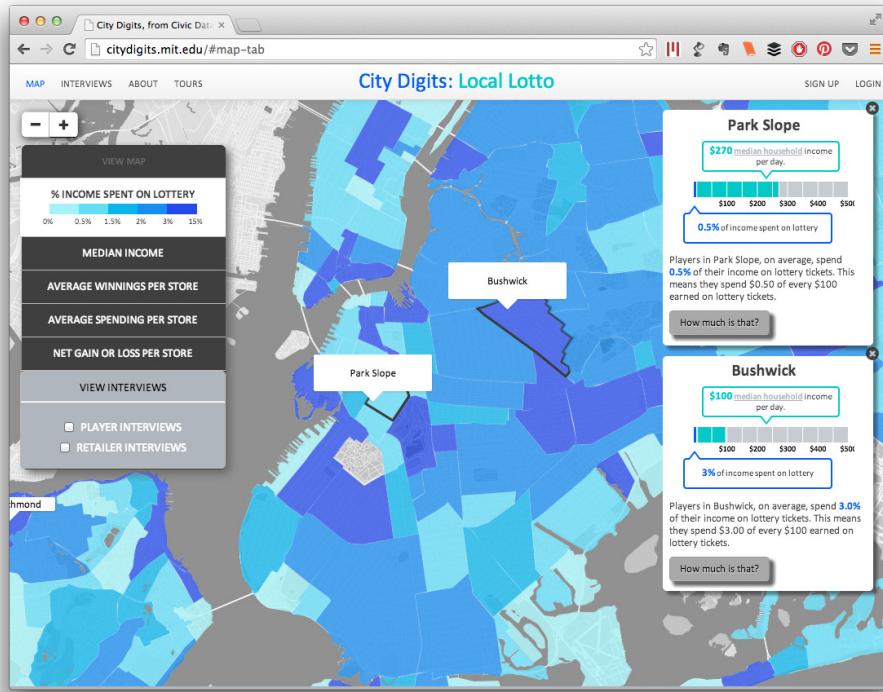
**Fig. 3.10** Average spending per store map: the darker green represents areas where, on average, people spent more on lottery tickets at each store



**Fig. 3.11** Net gain or loss map, zoomed into the street level: the green circles represent the amount of money spent on lottery tickets at individual retail stores; the purple circles represent the amount of money won from lottery tickets at the same stores



**Fig. 3.12** % income spent on lottery map: the darker blue represents areas where a greater percentage of median household income was spent on lottery tickets



## Constructing Data-Supported Arguments

In the final phase of the curriculum, students synthesize knowledge drawn from interview data and map data. Using the skills they developed and the information they collected during previous phases of the project, students combine qualitative and quantitative reasoning and use data to create an assessment of the lottery's social impact. During this activity, students are first presented with example statements about the lottery, such as "the lottery is a tax on the mathematically illiterate." The class debates these views and brainstorms new ideas before breaking into groups to strengthen a single argument, which they are asked to support using evidence from the class. Students share, debate, and publish their opinions about the lottery, creating multimedia narratives to illustrate their reflections using maps, interviews, and other evidence to support and structure their arguments.

Students expressed a diverse range of opinions in their work. For example, one group argued that “the lottery is a tax to manipulate and mislead the mathematically illiterate” since low-income communities spend so much on lottery tickets without getting significant return. The students used a map showing lottery winning and spending in Ridgewood to support their argument. “As you can see, it shows that in an average day in 2010, players will spend an approximate amount of \$737. Players will only receive \$40. Does this make sense? Is the Lottery fair?” They then used data collected from an individual interview as a concrete illustration of the same point. “Daquan lives in Bushwick, Brooklyn. Daquan is an older adult who, on average, spends 50 dollars a week participating in the lottery. However, the most he’s ever won was 100 dollars. He has yet to make up for the money that he’s already spent.” Another group disagreed with the statement that the lottery is a tax on the mathematically illiterate. Instead, they argued that the problem is that the lottery commission “omits essential information that can most likely persuade people not to play the lottery. For example many lottery players may not know that they lose 29 out of 30 times because that information isn’t displayed or highlighted.” One group proposed a solution for the inequality they perceived in the system: each borough should have its own lottery so that more of the money spent on lottery tickets would funnel directly into the community’s own education budget.

Another group of students made the opposite argument, that the lottery is good for the city, since the proceeds benefit the educational system and people are not forced to play. These students noted that 38% of lottery ticket proceeds go toward education and decided to use the maps to verify the significance of this amount. After reviewing different variables on the map, they shaped the argument that players at a single store in Woodside spend \$1647 every day on lottery spending and that 38% of that, or \$625.86, goes to schools. If you add up the proceeds for every store in the city for every day of the year, they argued, it becomes evident that the lottery is a significant resource for schools. Furthermore, the group

pointed to several interviews they conducted with neighbors who chose not to spend their money on the lottery because they knew they were unlikely to win. The students argued that these perspectives demonstrate that playing the lottery is a personal choice, so it should not be considered a tax on the poor.

### **3.2.3 Outcomes**

#### **Short-Term Outcomes: Data Literacy Competencies**

When we started the program, many of the students we worked with were unable to read a map, a problem we had not anticipated. We saw in the pre-pilot focus group that students did not have a sense of city geography, and that many could not locate the borough of Brooklyn on an unlabeled city map. This issue persisted even after reviewing the geography in class, an issue we will take into consideration when designing the next iteration of the curriculum and tool. Others had not mastered foundational mathematics skills on which the data analysis relied, such as calculating the median value of a set of numbers. We had to be flexible in the scope of the curriculum and improvise in order to support different levels of understanding.

Despite these challenges, the data explorations succeeded in engaging students, even those who did not normally participate in class. During the post-pilot focus group, the students told us they enjoyed the opportunity to explore data individually and in groups, rather than listening to a teacher lecture. One student said he had “never seen a class even work like this ever, ever before... Everyone was so interested. I didn’t think that was possible.” Students explained that they were engaged in the class because the issues were complex and “real” and demanded higher levels of reasoning than they had previously encountered.<sup>20</sup>

Exploring local data offered diverse opportunities for engagement, especially for English Language Learning students, who were not normally outgoing in class discussions.

Since the high school where we conducted the pilot was in a neighborhood that was heavily Spanish speaking, students who were native Spanish speakers were positioned to become leaders in their groups. One student who is a recent immigrant and was quiet during the initial class sessions became outgoing and animated during the data collection day. She conducted the majority of the interviews and was excited to contribute during the classroom interview debrief.

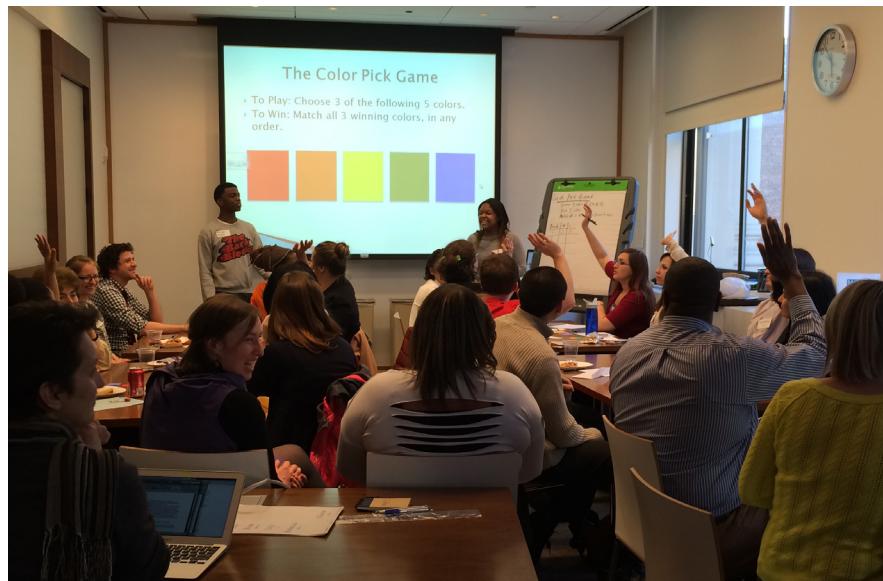
Students' data-supported arguments demonstrate that by the end of the course they were able to interpret data, develop their own arguments, and use a combination of quantitative and qualitative data to support them. Through both the lens of personal accounts from their neighbors and broader patterns of lottery spending habits by geography and demographic, students had to weigh the inequalities and benefits of the system. Analyzing these combined sources led students to develop strong opinions; by the time students developed their arguments, they felt like experts on the topic. There was a dramatic transformation between the first class and the last, as students became more confident speaking up and sharing their opinions in class. This motivation to share their opinions extended beyond the classroom—some students told us that they had discussed the topic with their friends and families. One student said he has been showing his mother his work and that he convinced her to stop buying lottery tickets.

### **Intermediate-Term Outcomes: Active Citizenship and Continued Engagement with Data**

As mentioned at the beginning of this chapter, the City Digits pilot was too recent and too small to have provided sufficient data to assess intermediate-term outcomes. However, the fact that the curriculum succeeded in deeply engaging a group that had previously been disengaged in math class—and that students developed such strong and sophisticated opinions on a complex topic—suggests promising possibilities for future engagement.

One of the most exciting outcomes in the months after the pilot is that five of the students—two from the initial pilot in May and three from the second-round pilot in November—volunteer to teach an abbreviated version of the class to groups of high school math teachers during the City Digits team’s training sessions for teachers interested in learning about the curriculum (Fig. 3.13). They have traveled to Los Angeles to teach the class during the Creating Balance Conference on Math Education and Social Justice and have lead the event multiple times at the Math for America offices in New York City. During these sessions, the students first lead the teachers through an exercise in calculating probability; then they lead the teachers through an analysis of the map data; finally they lead a discussion on the lottery’s impact. Most of these students were not the highest performers in their class during the pilot, however it is clear during these events that they have gone on to fully master the content they are teaching. They are articulate and enthusiastic presenters comfortable talking in front of a group, answering questions, and leading a discussion. While even the most hardworking students were occasionally disengaged in a school setting, here, the excitement of a higher stakes performance motivated them to step up to the plate.

**Fig. 3.13** Students lead teachers through an activity calculating probability at a teacher training event



During the latest event at the Math for America offices, the students asked the audience some difficult questions: “If most of the money [from the lottery] is going to education, how come many of our schools are in trouble of being shut down?”; they shared personal anecdotes: “My aunt plays \$200 every day, and she doesn’t care [that she isn’t winning]—she says her numbers are lucky and she has faith....If you have friends and family that play, how do you convince them [not to]?”; and one student told the audience that teaching City Digits gave her new confidence: “I doubted myself. I didn’t think this project would go this far. I want to be a public speaker now.” Through continued engagement outside the classroom, these students have not only deepened their knowledge of the topic but have also surpassed our expectations by embracing the opportunity to share their knowledge and perspective with others.

### **3.2.4 Conclusion**

In City Digits, we introduced students to working with data for the first time. These students will need continued support in order to deepen and apply their data literacy to other contexts. Yet we are hopeful that students’ success in this venture will help demystify the concept of data, making them more confident in their ability to interact with quantitative and qualitative data and more likely to seek new opportunities to work with data in the future. During the program, students began to see themselves as experts and educators. Over the long-term, we hope that the program will prepare students to be informed consumers of data, active civic participants, skilled critical thinkers across quantitative and qualitative boundaries, and also that some will go on to pursue more advanced levels of data science. We saw evidence in the teacher training sessions that City Digits helped students develop the confidence and motivation to become advocates and stakeholders in an issue. We hope this experience will also help them to become more active civic participants—and that their ability to support their arguments with data will enable their voices to be heard.

Implementing this curriculum was not easy. When we developed the curriculum, we relied on assumptions about what skills students would have already learned; some of these assumptions proved to be false and posed barriers to learning. In addition, the project faced numerous institutional constraints because it was implemented in a public school, which limited possibilities for the content and methods that could be used. Finally, the project required an enormous commitment by the curriculum development team and the teachers we worked with, a challenge that will make it difficult to bring this project to scale. In Chapter 4, “Data Literacy Challenges and Design Principles,” I will further discuss these challenges and propose strategies to address them in the future.

### **3.3 Conclusion**

In this chapter, I described two case studies that not only taught participants data literacy competencies but also laid the pathway for greater engagement in the future. Both programs witnessed forms of engagement that they could not have anticipated—for instance groups of students forming their own companies, or students becoming teachers and sharing their knowledge with new audiences. These unexpected long-term outcomes are crucial, because they represent an initiative’s greatest impact; yet they are more difficult to measure than short-term competencies. As we begin to witness new data literacy initiatives develop, it will be important for researchers to think widely about metrics in order to account for and bring attention to these surprising impacts.

## NOTES

- 1 Patrick Wintour, “Senior Cabinet Members Join Reform Push in Wake of MPs’ Expenses Scandal,” *The Guardian*, May 17, 2009, sec. Politics, <http://www.theguardian.com/politics/2009/may/18/cabinet-reform-mps-expenses-scandal>.
- 2 All quotes from Emma Mulqueeny are from a phone interview I conducted on January 17, 2014.
- 3 “Young Rewired State 2013: A Mentor’s Perspective | Blog,” *Box UK*, accessed March 16, 2014, <http://www.boxuk.com/blog/young-rewired-state-2013-mentors-perspective/>.
- 4 “YRS Festival of Code – The First 3 Days | Pearson Developer,” accessed March 16, 2014, <https://developer.pearson.com/blog/yrs-festival-of-code-the-first-3-days/>.
- 5 “Young Rewired State 2013.”
- 6 All Tweets with the hashtag #yrs2013 are archived at <http://eventifier.com/event/yrs2013/>.
- 7 “Young Rewired State 2013.”
- 8 *Code on the Road - Powered by Twilio*, 2013, [http://www.youtube.com/watch?v=kcZwh25SQIo&feature=youtube\\_gdata\\_player](http://www.youtube.com/watch?v=kcZwh25SQIo&feature=youtube_gdata_player).
- 9 “Young Rewired State 2013 at Manchester Digital Laboratory,” *YouTube*, accessed March 17, 2014, <http://www.youtube.com/watch?v=-rpBeD3QA7U>.
- 10 *Alex, 18 - Festival of Code Interview #yrs2013*, 2013, [http://www.youtube.com/watch?v=cepkLJpMW98&feature=youtube\\_gdata\\_player](http://www.youtube.com/watch?v=cepkLJpMW98&feature=youtube_gdata_player).
- 11 Parts of this case study draw from my contributions to the article “City Digits: Developing Socially-Grounded Data Literacy Using Digital Tools” co-authored with project collaborators Sarah Williams, Laurie Rubel, and Vivian Lim. Sarah Williams et al., “City Digits: Developing Socially-Grounded Data Literacy Using Digital Tools,” Unpublished Manuscript, 2014.
- 12 This material is based upon work supported by the National Science Foundation under Grant No. DRL-1222430. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.
- 13 Sarah Williams et al., “City Digits: Developing Socially-Grounded Data Literacy Using Digital Tools,” Unpublished Manuscript, 2014.
- 14 Al Baker and Motoko Rich, “New York City Students Show Slight Gains on Test Scores,” *The New York Times*, December 18, 2013, sec. Education, <http://www.nytimes.com/2013/12/19/education/new-york-city-students-show-slight-gains-on-test-scores.html>.

- 15 David A. Gruenewald, "The Best of Both Worlds: A Critical Pedagogy of Place," *Educational Researcher*, no. 4 (2003): 3, doi:10.2307/3700002; Roger Hart and Unicef., "Children's Participation," 1997; David Sobel, *Mapmaking with Children: Sense of Place Education for the Elementary Years*, 1998.
- 16 Gloria Ladson-Billings, "Toward a Theory of Culturally Relevant Pedagogy," *American Educational Research Journal*, no. 3 (1995): 465, doi:10.2307/1163320.
- 17 Vivian Lim et al., "Local Lotto: Mathematics and Mobile Technology to Study the Lottery," in *Cases on Technology Integration in Mathematics Education*, ed. D. Polly (Hershey, PA: IGI Global, 2014).
- 18 An advisory class is a daily mandatory period in which there is no set curriculum.
- 19 Williams et al., "City Digits: Developing Socially-Grounded Data Literacy Using Digital Tools."
- 20 Ibid.



## CHAPTER 4

# Data Literacy Challenges and Design Principles

The goal of this thesis has been to understand the learning and engagement that take place as youth learn to work with data and to identify a range of practices that support youth in the development of data literacy. Chapter 1 outlined the changing social factors that have made data a defining and problematic force in society, arguing that data literacy should be considered a critical set of skills and knowledge for young people. Chapter 2 introduced the concept of data literacy, described its objectives and methods, and surveyed a range of existing initiatives that further help define the field. Chapter 3 described two examples of data literacy initiatives—Young Rewired State’s Festival of Code, an out-of-school data hackathon for youth in the UK, and City Digits: Local Lotto, an in-school mathematics class centered around data collection and analysis. Through my research on data literacy initiatives in Chapters 2 and 3, I address two of my primary research questions: *How are initiatives in public schools and informal learning environments supporting youth in developing data literacy? What attributes characterize successful initiatives?*

In this concluding chapter I will now discuss some of the themes that unite the case studies and the field as a whole. By connecting these threads, I will address the questions that have guided this research: *What are the challenges and limitations of these initiatives?*

*What principles should guide the design and implementation of future data literacy initiatives and the digital technologies that support them?* I will analyze the challenges facing data literacy initiatives, propose strategies for overcoming these challenges, and define priorities for the design of future data literacy initiatives. This chapter is structured around three types of challenges, illustrated with evidence from the Festival of Code and City Digits case studies. From each type of challenge, I propose a corresponding design principle. First, I describe the challenges of implementing data literacy initiatives within American public school environments and provide suggestions for ways in which we can design an ecosystem for data literacy in both schools and informal learning environments. Next, I describe the challenge of diversity and access that data literacy initiatives across different types of learning environments face, and I describe strategies for engaging broader audiences. Finally, I discuss the pedagogical challenges of supporting data literacy in the classroom, including the lack of resources and training for teachers and the inherent difficulties of supporting personalized and customized learning in the classroom. I suggest that we need to create more training resources for teachers as well as design technologies to facilitate open-ended learning. Through this analysis, I propose an agenda of three design principles to guide researchers, educators, and practitioners working in this field.

**Fig. 4.1** Data Literacy Challenges, Examples, and Design Principles

Challenges	Examples	Design Principles
<b>Institutional Constraints of Public Schools</b>	<b>City Digits:</b> Common Core Standards limit student-centered learning; schools do not incentivize teachers to test new methods; short class periods constrain exploration	<b>Design Principle 1:</b> Create a Data Literacy Ecosystem In and Out of Schools
<b>Diversity and Access Challenges</b>	<b>Young Rewired State:</b> socioeconomic stereotypes of computer science field reproduced; resource intensive  <b>City Digits:</b> accessible only to students in participating classes; resource intensive	<b>Design Principle 2:</b> Design for a Range of Levels, Communities, and Learning Styles
<b>Pedagogical Challenges</b>	<b>Young Rewired State:</b> personalized learning model relies on large number of skilled mentors  <b>City Digits:</b> need for extensive teacher training; conflict between youth agency and understanding; customized content and technology difficult to replicate	<b>Design Principle 3:</b> Create Training Resources for Teachers and Support Open-Ended Learning

## 4.1 Institutional Constraints of Public Schools

Data literacy initiatives designed for public schools face many institutional constraints, because they must be tailored to fit the existing format of classroom learning. Any program too far outside traditional boundaries is unlikely to be adopted in a public school classroom. As a result, initiatives that hope to gain traction within public schools are forced to strike a balance between fitting into the current system and trying to innovate. Institutional

constraints range from strict curriculum requirements; to a lack of time, training, resources, and incentives for teachers; to a conflict between learning goals and the structure of the school day.

First, getting the support of the school administration typically requires demonstrating that a curriculum supports the Common Core State Standards, an education initiative in the United States that aims to make education standards consistent across states by setting objectives for what students in grades K-12 should learn in English language arts and mathematics in each grade. The Common Core State Standards have been adopted in 46 of 50 states, in part due to the federal financial incentives established by the Race to the Top program.<sup>1</sup> Designing a curriculum specifically to support learning of the standards limits the type of content covered, since standards- and non-standards focused content compete for limited class time. In the case of City Digits, the curriculum was designed to support both data literacy objectives and Common Core Mathematics objectives, a set that overlapped but was not a perfect match. Where they diverged, there was a tendency to emphasize Common Core Mathematics objectives, since those were the skills being assessed and reported. In particular, this meant striking a balance between quantitative reasoning, which aligned well with mathematics standards, and qualitative reasoning, which was not a specified Common Core Mathematics goal.

In addition, core requirements and high stakes testing in public schools lead to teaching styles that reduce the opportunity for personal exploration that is critical to developing data literacy. Freeform project-based learning models, such as the one implemented at the Festival of Code, are poorly suited to the current public school structure, because learning is difficult to assess and does not correspond to a standardized set of learning goals for all participants. If the City Digits curriculum had been designed without the standards in mind, it would have been possible to give students a more active role in determining the topic of inquiry; participating in the research, data analysis, and

map-making processes; and designing interventions based on their conclusions. Designing for the Common Core Standards meant the curriculum had to focus on narrower aspects of the data exploration process.

Furthermore, teachers are pushed to teach to the standards as efficiently as possible, a model which conflicts with deep data exploration. In a typical math classroom, this translates into teachers demonstrating concepts and asking students to memorize formulas and complete practice problems. By contrast, City Digits attempted to disrupt this model by giving students time to make discoveries on their own. Instead of teaching about isolated concepts, it emphasized contextualization and validated a wider spectrum of learning. This type of learning initiative is designed to facilitate personal exploration, not to optimize efficiency—an approach that often conflicts with schools' system of values.

Schools do not incentivize teachers to test new methods. As we found with City Digits, implementing an experimental curriculum can be a challenging and time-consuming process, both in terms of preparation and implementation. To prepare to teach a new initiative, a teacher must devote significant time outside the classroom to learning methods, content, and technology. The teachers who participated in the City Digits pilot not only devoted time to training; they also worked with the design team to refine the curriculum, tailored content to the needs of their own classes, and traveled across the country to speak about the work at conferences and teacher training sessions. Even after extensive training, implementation of the curriculum in the classroom can be unpredictable and difficult. One of the teachers who implemented the City Digits pilot described the curriculum as both a way to use a social justice issue to teach students about mathematics, and inversely, as a way to use mathematics to teach students to interpret their world—an approach that was challenging to teach and often frustrating for students, who were accustomed to being passive consumers of content. The City Digits team was fortunate to work with several teachers eager to donate their time to strengthening and implementing the curriculum.

But when the labor cost of implementing an initiative is borne by individual teachers rather than being supported by the parent institution, it is unrealistic to expect a large number of teachers to step up to bat. Not all teachers would want to adopt a challenging, time-consuming, and minimally vetted curriculum or would have the technical skills to facilitate the use of new technologies in their classroom.

A final challenge is the structure of the school day: short class periods make it difficult to reach a point of full saturation and engagement. During the Festival of Code, students work in sessions that span the entire day, during which they become deeply absorbed in complex tasks. During the City Digits curriculum, however, several of the activities felt rushed or were cut short by the end of the period, precluding the possibility of sustained engagement. The time-consuming nature of data exploration and analysis activities make them best suited to unhurried exploration.

### **Design Principle 1: Create a Data Literacy Ecosystem In and Out of Schools**

The challenges described above illustrate that there are limited opportunities to scale data literacy initiatives in public schools, and that the public school environment is rarely conducive to innovation. Spreading an initiative in schools is challenging whether advocates take a top-down approach, wooing school or district administrators, or a bottom-up approach, wooing individual teachers. Either way, they must both convince schools that an initiative supports core learning goals and convince teachers that an initiative is worth the effort. Curricula will at first have to be designed conservatively to work in these settings and may be limited in scale.

Despite these challenges, I believe it is important for researchers, educators, and practitioners to continue efforts to implement data literacy initiatives in schools. Even though data literacy initiatives designed for informal learning environments are subject to few institutional constraints, the problem with focusing on developing data literacy

initiatives out of schools is that many young people get left out. Introducing data literacy in the classroom is crucial for broadening participation, since it engages young people who do not have the opportunity to encounter data at home or in an extracurricular program.

To confront these challenges, we need short-term and long-term goals for getting data literacy into schools. In the short-term, this will mean providing teacher training resources and support for individual teachers, a strategy I discuss further in Section 4.3, “Pedagogical Challenges.” Integrating small-scale data literacy programs into schools will enable us to test a range of pedagogical methods in the school environment and educate teachers and administrators about data literacy. In order to sustain and spread data literacy over the long-term, data literacy should ultimately be integrated into the teacher certification process, which will require the support of teachers unions, higher education, and administrations at the district, state, and national levels.

To work toward this goal, we must start by spreading broader awareness about data literacy so that schools and the general public begin to recognize its importance. Out-of-school data literacy programs will play a major role in both creating greater awareness about data literacy and advancing innovative methods that could not be incubated within the classroom. Out-of-school programs serve as an unconstrained testing ground for disruptive data literacy pedagogy. Once new methods have been tested and refined outside schools, it becomes easier to introduce them into the classroom.

Ultimately, we need to establish an ecosystem for supporting data literacy both inside and outside schools. In-school programs will ensure that all students are given an opportunity to be exposed to data and to develop basic data literacy, while out-of-school programs will provide students who seek further engagement with a venue to pursue more advanced and independent data exploration. To do this effectively, we must rethink the dichotomy between in-school and out-of-school learning environments in order to create opportunities that are mutually supportive and that take advantage of the structure of each

respective learning environment. Although the school environment can seem constricting, it also offers inbuilt opportunities to structure and scaffold learning that can be beneficial, especially for young people who are being introduced to a topic for the first time. Informal learning, in turn, should build upon these experiences and encourage independence and freeform exploration. To build and expand this ecosystem, a critical challenge will be collaboration across practices and across learning infrastructures. Stakeholders designing and implementing data literacy initiatives will need to share pedagogical methods and build practices that reinforce each other.

As this ecosystem evolves, it will be important to develop better metrics for assessing the outcomes of data literacy initiatives. Since schools are obsessed with measurement, metrics will ease schools' resistance and help administrators see a program's effectiveness and value. This is a task that researchers and program facilitators will need to take on in order to strengthen and legitimize the discussion about data literacy and help data literacy initiatives fit into formal educational infrastructures. In addition, in order to create and sustain a robust data literacy ecosystem in and out of schools, it will be important to define indicators for progress within and across initiatives—beginning with measuring where we stand now, in order to have a point of comparison for future progress. These metrics will enable us to iterate and improve initiatives.

## 4.2 Diversity and Access Challenges

As we have seen with the Festival of Code and City Digits, it can be a challenge to engage diverse audiences from demographics that are traditionally underrepresented in the fields of programming and data science. As we examine the composition of the audiences that these initiatives engage, it is important to understand who is missing in those audiences and to evaluate how to engage populations who are excluded. Exacerbating this challenge

is the fact that creating a data literacy initiative is resource intensive. Not only are there considerable monetary costs, but these initiatives also rely on skilled personnel to devote considerable time to development and implementation. These difficulties leave underserved communities ill equipped to create and implement such initiatives.

Although Young Rewired State solved its initial audience dilemma—a lack of youth interested in programming—it has encountered a new issue, limited diversity of participants, in particular a lack of female participants. Since the initiative is aimed at youth who have some experience programming, it excludes those who lack access to technology or exposure to code. Gender biases that exist in the adult programming community are reflected and reinforced in the youth coding community. Even though Young Rewired State has worked to encourage more female participants, and has made accommodations to enable youth from low-income families to participate, these efforts are not enough to eliminate the existing socioeconomic barriers that prevent many youth from becoming interested in code in the first place.

In addition, implementing the Festival of Code was resource intensive, which implies that replicating such a program will be easier for communities and schools with more resources and better networks. The initiative relied heavily on funding from corporate sponsorship as well as a large pool of partner institutions and individual volunteers. Last year, the Festival's operating budget was approximately £100,000, which was allocated largely toward hosting the weekend showcase: renting the venue, covering meals and accommodations for the 600 attending participants, resources like air conditioning and additional Wi-Fi capacity, and event resources and staff. In 2014, the budget will rise to around £250,000. In addition to financial sponsorship, around 150 volunteers and around 400 mentors provided a week of their time without compensation. Mulqueeny was well positioned to grow Young Rewired State's operations because she had a ready supply of volunteer data scientists and was closely tied to the tech community. YRS illustrates that to

operate at a large scale you need a strong network of supporters.

One of the greatest benefits of hosting City Digits in a public school environment was that it was positioned to engage youth across these socioeconomic barriers. By introducing youth to data for the first time, City Digits had the opportunity to make youth curious about other ways of working with data, an interest that could lead to future extracurricular engagement. Of course, City Digits faces its own accessibility challenge—it is limited to participants who live within a specific geographic region and whose teachers choose to adopt the curriculum.

City Digits is a similarly resource-intensive initiative, funded by a three-year grant from the National Science Foundation, of which approximately \$250,000 was allocated for developing the Local Lotto curriculum. These resources were devoted primarily to the team's labor: research, development of the curriculum, design and programming for the website, teacher training, research conducted in the classroom during the two pilots, and the production of academic papers, presentations, and materials to aid teachers wishing to implement the curriculum. In addition, funding was allocated for purchasing a set of tablets used in the classroom during the two pilots. For schools without the budget for tablets, portions of the curriculum are inaccessible.

Since developing data literacy initiatives can be a resource-intensive process, there is a risk that this model of education mirrors and exacerbates existing social inequalities. Wealthier neighborhoods have an inherent advantage in pooling and advocating for the resources necessary, while poorer communities may lack both the intrinsic resources and experience promoting their initiatives. The networked nature of these initiatives also makes them liable to reproduce social inequality rather than address it. Their success depends on the construction of large collaborative networks across a wide range of fields and geographies. Sociologists have long recognized that well-resourced individuals are more likely to possess large, so-called “weak tie” networks, which they are able to leverage to

their advantage. Members of poorer communities, by contrast, tend to have social networks that are denser, but less widely distributed.<sup>2</sup> As a result, they are less likely to be in a strong position to build broad networks to support their activities. This is a pitfall that must be considered as these programs are developed.

### **Design Principle 2: Design for a Range of Levels, Communities, and Learning Styles**

To broaden the range of youth participating in these activities, it is especially important to design opportunities that are aimed at beginners and that proactively solicit participation from underserved and underrepresented communities. In addition, the success of organizations such as Black Girls Code, which aims to increase the number of women of color in computer science, illustrates that learning environments that bring together individuals who share a common cultural background can help close this gap. Black Girls Code founder Kimberly Bryant explains that one of her motivations for starting the organization was that she felt culturally isolated as a black woman pursuing a career in computer science.<sup>3</sup> By providing girls with the opportunity to acquire skills along with their peers and meet role models they can relate to, the organization helps build confidence and community.

Another strategy to engage a broader range of youth is to specifically create opportunities for youth who learn in different ways. One example of this is illustrated by the online community that has formed around Scratch. On the Scratch site, there are categories for animations, art, games, music, and stories. By encouraging a diverse range of endeavors, Scratch becomes accessible and appealing to young people who are not inherently drawn to programming, especially those who do not tend toward a logical-mathematical learning style, but who have other interests like drawing or music.<sup>4</sup> This helps close the gap by lowering the barriers to participation for beginners, since youth can utilize skills that they already feel confident about. It also gives youth who believe they are not good at math or

computer science because of how it has been taught to them in the past a fresh opportunity at engagement using methods more suited to their style of learning. Data literacy programs can borrow from this model by making explicit the wide range of possible methods, topics, and products that data exploration encapsulates and by giving youth the option to engage with data in different ways. While some individuals might be drawn to building electronics that tell them about the world around them, others might prefer to express their findings through visual art, or find motivation in investigating issues affecting their communities. One implication of this insight is that the language and framing around these activities should be carefully articulated to communicate their inclusivity.

Finally, to address the resource challenges that make data literacy initiatives inaccessible for poor communities, it is critical to design resources that are easily replicable. Institutions such as universities, corporations, and foundations are already taking on the high costs of developing, testing, and refining new methods and technologies for data literacy. If these methods and technologies are free and publicly available, schools and community groups with fewer resources will be able to implement data literacy initiatives at a relatively low cost.

## 4.3 Pedagogical Challenges

Finally, teaching data literacy initiatives is challenging because it requires teachers to adopt new content, methods, and technologies. Most teachers themselves are not data literate, and many are uncomfortable using new technologies in the classroom. They lack sufficient training opportunities in this field, and they lack a community of support to assist them in strengthening their classroom approach. Right now, there are few existing data literacy initiatives to serve as models, let alone mechanisms for getting teachers interested in data literacy curricula, or resources for supporting teachers in learning new methods. With

Young Rewired State, subject-matter training was unnecessary because mentors were already experts; but most initiatives will not have access to a ready supply of volunteer data scientists. With City Digits, teachers needed to participate in several training sessions, led by the researchers developing the project, in order to become comfortable with new content, methods, and technologies. As we begin to design data literacy initiatives intended to be implemented on a larger scale, we will need to develop a sustainable model for teacher training and support.

Even for teachers who know the subject matter well, data literacy initiatives are inherently challenging to facilitate. Teachers must simultaneously assist students in learning complex material while also encouraging their personal agency through independent exploration and creation. This balance can be difficult to get right in the classroom—especially in schools, where independent learning is often in conflict with pedagogical norms, large class sizes limit possibilities for giving students individual attention, and time constraints restrict exploration. To pull this off, teachers must be able to tailor learning to audiences that represent a diversity of levels, interests, and needs. Learning must be personalized for individual students or customized for specific communities. Data literacy initiatives that embody these qualities are difficult to grow and replicate, since they require skilled facilitators, or significant labor to rework content for other contexts.

At the Festival of Code, there is no common learning agenda across participant groups, since youth develop their own project ideas. Participants represent a wide range of ages and levels, so learning must be customized to the level and needs of each participant. In many cases this translated into mentors directly instructing inexperienced participants, while experienced participants worked independently and occasionally asked for advice. This model relies on the availability of knowledgeable mentors who have multiple areas of expertise and are equipped to help youth acquire skills required for the particular projects they want to develop. There must also be a high ratio of mentors to participants—last year

the Festival of Code had one volunteer mentor for every three participants. Finally, a key characteristic is that youth of different ages and levels work together, so youth also serve as teachers and guides for their peers and contribute to an online support network for the larger community. It would only be possible to replicate this model of highly personalized customization given a large number of skilled mentors.

In City Digits, students were given more rigidly formatted opportunities for personal agency and exploration through the neighborhood investigations, map analysis, and reflection assignment. During these activities, teachers attempted to both guide students and allow them to work on their own, a balance that proved challenging given different levels of student engagement and understanding. During the interviews, it was simple to give students freedom and autonomy, because the skills required were intuitive to them. Yet during the data analysis activity, teachers had to integrate lectures in order to teach Common Core Standards and help students with poor map-reading skills learn to interpret the patterns they saw. The teachers found that explaining the complexities of the maps greatly diminished the time they were able to devote to independent exploration. In addition to the implementation challenges related to independent learning, City Digits faces a replication challenge due to the highly customized nature of its content, which is focused on a specific topic and only applicable to a single city. The long-term model for replicating City Digits is to develop new iterations around other current social justice topics, such as the new module the City Digits team is currently developing around pawnshops and fringe banking in New York City. Yet creating these new iterations requires developing new content and making significant modifications to the customized web platform, processes which are both labor intensive.

### **Design Principle 3: Create Training Resources for Teachers and Support Open-Ended Learning**

In order to acquire data skills, learn to effectively facilitate student-driven learning, and find advice as they tailor content for diverse audiences, educators need access to training resources and community support. They should have ready access to information and curriculum guides, in-person training sessions, and opportunities to interact with other educators who are tackling similar issues in their own classrooms. One example of a rich repository of educator resources is the ScratchED website for educators using Scratch in their classrooms. The ScratchED website provides curriculum guides; allows teachers to share stories, resources, and discussions about designing learning environments; and facilitates online and in-person gatherings for teachers, including workshops, meet-ups, and webinars.<sup>5</sup> Having this type of connection point to discuss strategies for facilitating data literacy in the classroom transforms an isolated challenge into a communal one. We need more spaces where educators can share knowledge about the challenges of supporting data literacy and begin to form a community around data.

Another way to address the challenge of supporting student-driven learning is through the development of better open-ended technologies for data exploration. Currently initiatives like the Festival of Code, where youth get hands-on experience with all stages of data exploration, are only accessible for communities that have access to facilitators with significant experience in data science and youth who have some background in computer science. Recent technologies have begun to make data exploration more accessible to the general public, but many still require an understanding of computer science or lack sufficient resources to be used widely in educational settings. Software like Tableau, Excel, and Google Fusion Tables, for instance, start to do this by enabling users to upload data and create simple data visualizations without knowledge of programming. The recent “Cloud data-structures” addition to Scratch goes further by enabling youth to work with data within the context of their interactive Scratch creations. “Thus, a young programmer who

is interested in creating a game gets to explore storing, sorting, and retrieving online data as a part of building a high-score list for her game with Cloud data-structures. Similarly, a programmer interested in storytelling gets to explore storing and retrieval of online data by using Cloud data-structures to keep track of user-choices for his choose-your-own-adventure story.”<sup>6</sup> The development of tools to lower barriers to entry of comparing, analyzing, and visualizing data, as well as resources to guide their use in educational contexts, will make data exploration accessible to broader audiences and will enable open-ended learning.

In addition, in order to address the issue of customization, educators need technologies that make it easier for them to develop their own content using data. This is something the City Digits team hopes to implement in the future in order to enable educators to use the City Digits model without being tied to its current specificities. Such a tool would make it simple to collect local data, analyze quantitative data, and discuss the patterns revealed—for any location, around any locally relevant topic. Currently, we have seen instances where teachers have used Social Explorer or Google Earth to create simple maps for data analysis. Yet these methods are not accessible for most educators, who, like their students, lack data literacy.

Despite the challenges of personalization and customization, I believe it is critical to resist the urge to standardize that so often accompanies attempts to achieve greater scale. Initiatives should customize content and personalize learning in order to deeply engage participants. This means that initiatives might not necessarily be appropriate for all contexts—they will succeed only if the teachers or facilitators are prepared, interested, and have access to the resources they need. In order for initiatives to grow sustainably, adoption should always be what educators choose, not something forced on them by the administration. Furthermore, the same lessons might succeed with one group of youth only to fall flat with another because of their level of background knowledge, interests,

or learning style. It is critical for facilitators to know their audience and be able to select strategies and adjust content accordingly.

## 4.4 Conclusion

In this chapter, I have proposed three principles to guide researchers, educators, and practitioners as they design future data literacy initiatives:

1. Create a data literacy ecosystem in and out of schools.
2. Design for a range of levels, communities, and learning styles.
3. Create training resources for teachers and support open-ended learning.

If we follow these principles, I believe we can make data literacy accessible to a broad range of communities through formal and informal learning environments. These principles are intended to serve as loose guidelines as data literacy becomes more defined as a social need and as a pedagogical pursuit. They will require further adjustment to meet the changing uses of data and the public's shifting understanding and level of data literacy.

This thesis has offered a road map for an emerging field—what it looks like now, where it is heading, and how we should help guide its development. It was inspired by missing links that I saw among current conversations around the ethical dilemmas surrounding data; the benefits of open data and participatory data collection; youth digital literacies; and the pedagogical practices enabled by digital media.

Instead of assuming that data exploration is an adult domain, we must open these practices to the next generation of learners. My research across different domains and learning environments suggests that we are at an inflection point—a moment at which youth data literacy has become an achievable goal. By supporting data literacy, we can empower young people to become experts of data, rather than its passive subjects. Data literate youth will understand how governments and corporations are using their personal

data, so they will be more equipped to control what data they release. They will understand data modeling and its ethical dilemmas, so they will be prepared to understand and question institutional decisions based on data. And they will have the tools to think critically and ethically about data as they grow up to become the data scientists of the future. I believe we have an opportunity to prepare a new generation to better understand the issues inherent in data analysis and learn responsible data practices, which will help prevent them from making the mistakes so prevalent in the history of data collection, analysis, and usage.

I hope to see a community of practice develop around building youth data literacy. Data literacy is not just a skillset; it is a stake in our communities of the future. We know from the cases I have described and others that there is a great opportunity to build data literacy and support young people in becoming engaged, articulate members of their communities. To do that, we need to design tools and resources to support data literacy. That is the challenge we as researchers, educators, and practitioners face.

## NOTES

- 1 “Understanding the Common Core Standards,” *Education Digest* 79, no. 8 (April 2014): 16.
- 2 Mark Granovetter, “The Strength of Weak Ties,” *American Journal of Sociology* 78, no. 6 (1973): 1.
- 3 “Black Girls Code” *Imagine. Build. Create.*, accessed April 11, 2014, <http://www.blackgirlscode.com/about-bgc.html>.
- 4 Mitchel Resnick et al., “Scratch: Programming for All,” *Communications of the ACM* 52, no. 11 (November 2009): 64.
- 5 Karen A. Brennan, “Best of Both Worlds: Issues of Structure and Agency in Computational Creation, in and out of School” (Massachusetts Institute of Technology, 2013), 45.
- 6 Sayamindu Dasgupta, “From Surveys to Collaborative Art: Enabling Children to Program with Online Data,” in *Proceedings of the 12th International Conference on Interaction Design and Children* (ACM, 2013), 29, <http://dl.acm.org/citation.cfm?id=2485784>.

# Bibliography

- “About NCHS - NCHS Fact Sheets - Overview.” Accessed May 5, 2014. [http://www.cdc.gov/nchs/data/factsheets/factsheet\\_overview.htm](http://www.cdc.gov/nchs/data/factsheets/factsheet_overview.htm).
- Alex, 18 - Festival of Code Interview #yrs2013*, 2013. [http://www.youtube.com/watch?v=cepkLJpMW98&feature=youtube\\_gdata\\_player](http://www.youtube.com/watch?v=cepkLJpMW98&feature=youtube_gdata_player).
- Baker, Al, and Motoko Rich. “New York City Students Show Slight Gains on Test Scores.” *The New York Times*, December 18, 2013, sec. Education. <http://www.nytimes.com/2013/12/19/education/new-york-city-students-show-slight-gains-on-test-scores.html>.
- “Big Data: The next Frontier for Innovation, Competition, and Productivity | McKinsey & Company.” Accessed February 16, 2014. [http://www.mckinsey.com/insights/business\\_technology/big\\_data\\_the\\_next\\_frontier\\_for\\_innovation](http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation).
- “Black Girls Code, BlackGirlsCode.” *Black Girls Code Imagine. Build. Create.* Accessed April 11, 2014. <http://www.blackgirlscode.com/about-bgc.html>.
- Brennan, Karen A. “Best of Both Worlds: Issues of Structure and Agency in Computational Creation, in and out of School.” Massachusetts Institute of Technology, 2013.
- Brown, Greg. “Why Kenya’s Open Data Portal Is Failing - And Why It Can Still Succeed.” *Sunlight Foundation*. Accessed December 7, 2013. <http://sunlightfoundation.com/blog/2013/09/23/why-kenyas-open-data-portal-is-failing-and-why-it-can-still-succeed/>.
- Buckingham, David. “Digital Media Literacies: Rethinking Media Education in the Age of the Internet.” *Research in Comparative and International Education* 2, no. 1 (January 1, 2007): 43–55.
- Butler, Declan. “When Google Got Flu Wrong.” *Nature* 494, no. 7436 (February 13, 2013): 155–56. doi:10.1038/494155a.
- Calzada Prado, Javier, and Miguel Ángel Marzal. “Incorporating Data Literacy into Information Literacy Programs: Core Competencies and Contents.” *Libri: International Journal of Libraries & Information Services* 63, no. 2 (June 2013): 123–34. doi:10.1515/libri-2013-0010.

- Carlson, Jacob, Michael Fosmire, C. C. Miller, and Megan Sapp Nelson. "Determining Data Information Literacy Needs: A Study of Students and Research Faculty." *Portal: Libraries & the Academy* 11, no. 2 (April 2011): 629.
- "China Prepares for Smart City Construction Boom." Accessed December 13, 2013. [http://www.chinadaily.com.cn/business/2013-10/16/content\\_17035874.htm](http://www.chinadaily.com.cn/business/2013-10/16/content_17035874.htm).
- Code on the Road - Powered by Twilio*, 2013. [http://www.youtube.com/watch?v=kcZwh25SQIo&feature=youtube\\_gdata\\_player](http://www.youtube.com/watch?v=kcZwh25SQIo&feature=youtube_gdata_player).
- Crawford, Kate. "Think Again: Big Data." *Foreign Policy*, May 10, 2013. [http://www.foreignpolicy.com/articles/2013/05/09/think\\_again\\_big\\_data](http://www.foreignpolicy.com/articles/2013/05/09/think_again_big_data).
- Crawford, Kate, Gustavo Faleiros, Amy Luers, Patrick Meier, Claudia Perlich, and Jer Thorp. *Big Data, Communities and Ethical Resilience: A Framework for Action*. Community Resilience through Big Data and Technology. Bellagio Center: Rockefeller Foundation, 2013.
- "Crowdsourcing Iceland's Constitution." *IHT Rendezvous*. Accessed December 12, 2013. <http://rendezvous.blogs.nytimes.com/2012/10/24/crowdsourcing-icelands-constitution/>.
- Darby, H. C. *Domesday England*. Cambridge University Press, 1986.
- Dasgupta, Sayamindu. "From Surveys to Collaborative Art: Enabling Children to Program with Online Data." In *Proceedings of the 12th International Conference on Interaction Design and Children*, 28–35. ACM, 2013. <http://dl.acm.org/citation.cfm?id=2485784>.
- diSessa, Andrea. *Changing Minds: Computers, Learning, and Literacy*. A Bradford Book, 2001.
- Flood, Joe. *The Fires: How a Computer Formula Burned down New York City—and Determined the Future of American Cities*. New York: Riverhead Books, 2010.
- "Gapminder." Accessed April 30, 2014. <https://sites.google.com/a/nycischool.org/gapminder/about-gapminder-at-the-ischool>.
- Gasser, Urs, Momin Malik, Sandra Cortesi, and Meredith Beaton. *Mapping Approaches to News Literacy Curriculum Development: A Navigation Aid*. SSRN Scholarly Paper. Rochester, NY: Social Science Research Network, November 14, 2013. <http://papers.ssrn.com/abstract=2354500>.
- Granovetter, Mark. "The Strength of Weak Ties." *American Journal of Sociology* 78, no. 6 (1973): 1.
- Greenfield, Adam, and Nurri Kim. *Against the Smart City (The City Is Here for You to Use)*. 1.3 edition. Do projects, 2013.

- Gruenewald, David A. "The Best of Both Worlds: A Critical Pedagogy of Place." *Educational Researcher*, no. 4 (2003): 3. doi:10.2307/3700002.
- Gurstein, Michael B. "Open Data: Empowering the Empowered or Effective Data Use for Everyone?" *First Monday* 16, no. 2 (January 23, 2011). <http://journals.uic.edu/ojs/index.php/fm/article/view/3316>.
- Hacker, J. David. "New Estimates of Census Coverage in the United States, 1850-1930." *Social Science History* 37, no. 1 (Spring 2013): 71–101. doi:10.1215/01455532-1958172.
- Hart, Roger, and Unicef. "Children's Participation," 1997.
- "IBM and City of Portland Collaborate to Build a Smarter City." CTB10, August 9, 2011. <http://www-03.ibm.com/press/us/en/pressrelease/35206.wss>.
- "Iceland: From the Financial Crisis to Open Data." *Open Knowledge Foundation Blog*. Accessed December 12, 2013. <http://blog.okfn.org/2010/03/26/iceland-from-the-financial-crisis-to-open-data/>.
- Janssen, Kathleen. *Open Government Data: Right to Information 2.0 or Its Rollback Version?* Rochester, NY: Social Science Research Network, September 1, 2012. <http://papers.ssrn.com/abstract=2152566>.
- Kehoe, Michael, Michael Cosgrove, S. D. Gennaro, Colin Harrison, Wim Harthoorn, John Hogan, John Meegan, Pam Nesbitt, and Christina Peters. "Smarter Cities Series: A Foundation for Understanding IBM Smarter Cities." *An IBM Redguide Publication*, 2011. <http://www.redbooks.ibm.com/redpapers/pdfs/redp4733.pdf>.
- Kitchin, Rob. "Four Critiques of Open Data Initiatives." *Impact of Social Sciences*, November 27, 2013. <http://blogs.lse.ac.uk/impactofsocialsciences/2013/11/27/four-critiques-of-open-data-initiatives/>.
- Ladson-Billings, Gloria. "Toward a Theory of Culturally Relevant Pedagogy." *American Educational Research Journal*, no. 3 (1995): 465. doi:10.2307/1163320.
- Lim, Vivian, Erica Deahl, Laurie Rubel, and Sarah Williams. "Local Lotto: Mathematics and Mobile Technology to Study the Lottery." In *Cases on Technology Integration in Mathematics Education*, edited by D. Polly. Hershey, PA: IGI Global, 2014.
- Livingstone, Sonia, Elizabeth Van Couvering, and Nancy Thumin. "Converging Traditions of Research on Media and Information Literacies: Disciplinary, Critical, and Methodological Issues." In *Handbook of Research on New Literacies*. New York: Lawrence Erlbaum Associates/Taylor & Francis Group, 2008.
- Manovich, Lev. "What Is Visualisation?" *Visual Studies* 26, no. 1 (March 2011): 36–49. doi:10.1080/1472586X.2011.548488.

- O'Connell, Pamela Licalzi. "Korea's High-Tech Utopia, Where Everything Is Observed." *New York Times* 5 (2005). <http://tofugalaxy.com/wp-content/uploads/2010/11/nytimes.pdf>.
- Resnick, Mitchel, John Maloney, Andrés Monroy-Hernández, Natalie Rusk, Evelyn Eastmond, Karen Brennan, Amon Millner, et al. "Scratch: Programming for All." *Communications of the ACM* 52, no. 11 (November 2009): 60–67.
- Singer, Natasha. "I.B.M. Takes 'Smarter Cities' Concept to Rio de Janeiro." *The New York Times*, March 3, 2012, sec. Business Day. <http://www.nytimes.com/2012/03/04/business/ibm-takes-smarter-cities-concept-to-rio-de-janeiro.html>.
- Sobel, David. *Mapmaking with Children: Sense of Place Education for the Elementary Years*, 1998.
- "Standards for Education Data Collection and Reporting (SEDCAR)," December 31, 1991. <http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=92022>.
- Street, Brian V. *Social Literacies: Critical Approaches to Literacy in Development, Ethnography and Education*. Routledge, 1995.
- Taylor, Nigel. *Urban Planning Theory Since 1945*. SAGE, 1998.
- Townsend, Anthony M. *Smart Cities : Big Data, Civic Hackers, and the Quest for a New Utopia*. New York: W.W. Norton & Company, Inc., 2013.
- "Understanding the Common Core Standards." *Education Digest* 79, no. 8 (April 2014): 16–21.
- US Census Bureau, Census History Staff. "1790 Overview - History - U.S. Census Bureau." Accessed December 11, 2013. [http://www.census.gov/history/www/through\\_the\\_decades/overview/1790.html](http://www.census.gov/history/www/through_the_decades/overview/1790.html).
- . "Counting the Population - History - U.S. Census Bureau." Accessed December 11, 2013. [http://www.census.gov/history/www/innovations/data\\_collection/counting\\_the\\_population.html](http://www.census.gov/history/www/innovations/data_collection/counting_the_population.html).
- Us, Line Industries | Terms of Use | Privacy Policy | Cookie Policy | Contact. "YRS Festival of Code – The First 3 Days | Pearson Developer." Accessed March 16, 2014. <https://developer.pearson.com/blog/yrs-festival-of-code-the-first-3-days/>.
- Williams, Sarah, Erica Deahl, Laurie Rubel, and Vivian Lim. "City Digits: Developing Socially-Grounded Data Literacy Using Digital Tools." Unpublished Manuscript, 2014.
- Willis, James J., Stephen D. Mastrofski, and David Weisburd. "Making Sense of COMPSTAT: A Theory-Based Analysis of Organizational Change in Three Police Departments." *Law & Society Review* 41, no. 1 (March 1, 2007): 147–88. doi:10.1111/j.1540-5893.2007.00294.x.

- Wintour, Patrick. "Senior Cabinet Members Join Reform Push in Wake of MPs' Expenses Scandal." *The Guardian*, May 17, 2009, sec. Politics. <http://www.theguardian.com/politics/2009/may/18/cabinet-reform-mps-expenses-scandal>.
- "Young Rewired State 2013 at Manchester Digital Laboratory." *YouTube*. Accessed March 17, 2014. <http://www.youtube.com/watch?v=-rpBeD3QA7U>.
- "Young Rewired State 2013: A Mentor's Perspective | Blog." *Box UK*. Accessed March 16, 2014. <http://www.boxuk.com/blog/young-rewired-state-2013-mentors-perspective/>.
- Zak, 8 - *Festival of Code Interview #yrs2013*, 2013. [http://www.youtube.com/watch?v=LKy2O0Ef5lI&feature=youtube\\_gdata\\_player](http://www.youtube.com/watch?v=LKy2O0Ef5lI&feature=youtube_gdata_player).