Interactive Data Narrative: Designing for Public Engagement

by

Heather Craig

Submitted to the Department of 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 2015

© Heather Craig, 2015. 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.

Signature redacted

Author........................................
Department of Comparative Media Studies
May 8, 2015

Signature redacted

Certified by....................................

Ethan Zuckerman
Director, Center for Civic Media
Thesis Supervisor

Signature redacted

Accepted by....................................

T.L. Taylor
Director of Graduate Studies, Comparative Media Studies
Interactive Data Narrative: Designing for Public Engagement

by

Heather Craig

Submitted to the Department of Comparative Media Studies on May 8, 2015, in partial fulfillment of the requirements for the degree of Master of Science in Comparative Media Studies

Abstract

Interactive data narrative, or the crafting of interactive online stories based upon new or existing data, has grown dramatically over the last several years. Data is increasingly available through such mechanisms as embedded sensor networks, remote sensing, and mobile data collection platforms. The affordances of mobile computing and increasing internet access enable widespread—and often citizen-powered—data collection initiatives. This proliferation of data raises the challenge of translating data into compelling and actionable stories. New data collection and online storytelling strategies foster a mode of communication that can reveal complexities, time-based shifts, and arcane patterns with regard to newly available geolocated data. This thesis investigates interactive storytelling as a mode of communicating data and analyzes trends and opportunities for future innovation. Surveying the field and analyzing specific projects lays the foundation for a design intervention for adding a narrative layer to geolocated, citizen-collected data.

Thesis Supervisor: Ethan Zuckerman
Title: Director, Center for Civic Media
Acknowledgements

First, thank you to my advisor, Ethan Zuckerman, for providing endless support and guidance and for challenging and encouraging me to pursue a project-based thesis.

To my committee, William Uricchio and Sarah Williams, for continual advice, feedback, and thorough reading of my draft.

To everyone who was willing to be interviewed for my thesis – these conversations were one of the most enjoyable aspects of working on this thesis.

To everyone at Comparative Media Studies and the Center for Civic Media for being an amazing community.

To Jim Paradis for the hiking suggestions when they were much needed.

To Sean Flynn, for always thinking the same thoughts, for the thesis co-working sessions, and for the collaboration.

To Dalia Othman, Ivan Sigal, and the Networked Storytelling group at the Berkman Center for the thoughtful conversation and inspiration for much of this work.

To the Datalore team – organizing and running the hackathon was one of the best experiences of grad school.

To my CMS class for making this experience especially meaningful – Chelsea Barabas, Sean Flynn, Desi Gonzalez, Suruchi Dumpuwar, Jesse Selle, Erik Stayton, Ainsley Sutherland, Liam Andrew, and Wang Yu.

To Jude Mwenda, for collaboration on the storytelling tool, and to Yu Wang for technical advice along the way.
# Contents

1 **Introduction** ................................................................................................. 7  
   1.1 Research questions and methodology .......................................................... 9  
   1.2 The challenge of defining data ..................................................................... 10  
   1.3 Chapter overview ...................................................................................... 11  

2 **History, theory, and motivation** ................................................................. 14  
   2.1 Historical precedent ................................................................................... 15  
   2.2 Proliferation of data .................................................................................. 18  
      2.2.1 Data collection tools ...................................................................... 19  
      2.2.2 Open data ....................................................................................... 24  
      2.2.3 Geolocated data ............................................................................. 25  
   2.3 Interactive narrative based on geolocated data ........................................... 25  
   2.4 Public engagement with interactive data narrative ..................................... 28  

3 **Design patterns and structural strategies** ..................................................... 30  
   3.1 Overview of projects .................................................................................. 31  
   3.2 Structural strategies ................................................................................... 32  
      3.2.1 Guided tour ...................................................................................... 33  
      3.2.2 Scrolling story ................................................................................. 34  
      3.2.3 Interactive web documentary ......................................................... 39  
   3.3 Data sources .............................................................................................. 44  
      3.3.1 Open data ...................................................................................... 46  
      3.3.2 Citizen-collected data ..................................................................... 48  
      3.3.3 Field data ...................................................................................... 51  
   3.5 Insights and opportunities based on field survey ......................................... 53  
      3.5.1 Greater integration of geolocated data with other content ............... 53
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.5.2 Highly guided user experience</td>
<td>54</td>
</tr>
<tr>
<td>3.5.3 Open development processes</td>
<td>55</td>
</tr>
<tr>
<td>3.5.4 Responsive story output</td>
<td>56</td>
</tr>
<tr>
<td>3.5.5 Lower technical and financial barrier to entry</td>
<td>57</td>
</tr>
<tr>
<td>3.5 Foundation for CartoStory</td>
<td>58</td>
</tr>
<tr>
<td>4 CartoStory: Design intervention</td>
<td>59</td>
</tr>
<tr>
<td>4.1 Impetus</td>
<td>60</td>
</tr>
<tr>
<td>4.2 Design process</td>
<td>61</td>
</tr>
<tr>
<td>4.2.1 Comparative analysis</td>
<td>64</td>
</tr>
<tr>
<td>4.2.2 Design and functionality</td>
<td>69</td>
</tr>
<tr>
<td>4.2.3 Differentiators</td>
<td>73</td>
</tr>
<tr>
<td>4.2.4 Technical implementation and initial prototyping</td>
<td>75</td>
</tr>
<tr>
<td>4.2.5 User types and scenarios</td>
<td>78</td>
</tr>
<tr>
<td>4.3 Future modifications</td>
<td>82</td>
</tr>
<tr>
<td>5 Conclusions and future work</td>
<td>83</td>
</tr>
<tr>
<td>5.1 Implications of innovation in interactive data narrative</td>
<td>84</td>
</tr>
<tr>
<td>5.1.1 Advocacy and intentionality</td>
<td>84</td>
</tr>
<tr>
<td>5.1.2 Public engagement with complex data</td>
<td>85</td>
</tr>
<tr>
<td>5.2 Future opportunities</td>
<td>87</td>
</tr>
<tr>
<td>5.2.1 Streaming real time data</td>
<td>87</td>
</tr>
<tr>
<td>5.2.2 Physical user engagement</td>
<td>90</td>
</tr>
<tr>
<td>5.2.3 Balancing customizability and technical accessibility</td>
<td>90</td>
</tr>
<tr>
<td>5.3 Conclusion</td>
<td>91</td>
</tr>
<tr>
<td>Appendix: Projects included in field survey</td>
<td>94</td>
</tr>
</tbody>
</table>
List of Figures

2.1 Page in 1874 *Statistical Atlas of the United States* ................................................ 17
2.2 Image from ProPublica’s *Losing Ground* ................................................................. 17
2.3 An early Ushahidi map .............................................................................................. 21
2.4 The Open Water Project’s Riffle .............................................................................. 23
3.1 The first interactive map in *Louisiana’s Moonshot* .............................................. 38
3.2 The intro to the “People” chapter in *Planet Money Makes a T-Shirt* .................. 42
3.3 A data visualization in the “People” chapter in *Planet Money Makes a T-Shirt* .... 43
4.1 Flowchart illustrating initial phases of the design process .................................. 64
4.2 The NetStories online tool database .................................................................. 65
4.3 An example of individual tool display on NetStories .......................................... 66
4.4 Initial design mockup ............................................................................................. 72
4.5 Example design of story output ............................................................................. 74
4.6 Early front-end of CartoStory created using Bootstrap with map base layer (Stamen) using Leaflet.js as a web mapping service .............................................. 77
4.7 Early version of CartoStory login page ................................................................. 79
4.8 User flow for two potential user types: citizen data collectors (user flow 1) or journalists and independent producers (user flow 2) ............................................... 81
Chapter 1

Introduction

Interactive data narrative, or the crafting of interactive online stories based upon new or existing datasets, has grown dramatically over the last several years. The rise of interactive data storytelling derives from the combined forces of proliferation of data, affordances of web technologies, and shifting audience expectations around content consumption. Data is increasingly available through such mechanisms as embedded sensor networks, remote sensing\(^1\), and mobile data collection platforms. The affordances of mobile computing and increasing internet access enable widespread—and often citizen-powered—data collection initiatives. This proliferation of data\(^2\) raises the challenge of translating data into compelling and actionable stories. While “big data” opens avenues for deeper understanding of issues like coastal erosion or urban infrastructure, “big data is not an ice-cold world of algorithms and automatons. There is an essential role for people, with all our foibles, misperceptions and mistakes, since these traits walk hand in hand with human creativity, instinct, and genius” (Mayer-Schonberger and Cukier, 2013, 196). The “essential role” for individuals and organizations includes translating data into narrative formats.

Major news outlets and independent interactive producers alike are grappling with the challenge of translating data into interactive online stories. Concomitant with the proliferation of

---

\(^1\) Remote sensing refers to aerial imagery such as satellite imagery.

\(^2\) This thesis employs the grammatical usage of “data” as singular or plural, rather than using the singular “datum” and plural “data.”
data procurement and collection initiatives, innovators in interactive storytelling produce increasingly beautiful and user-friendly interactive online content. Annotated maps and static data visualizations are being replaced by immersive interactive storytelling techniques, ranging from long-form scrolling stories to animated interactive visualizations. Organizations like The New York Times, National Film Board of Canada, NPR Visuals, and The Guardian increasingly emphasize data journalism and interactive content. Since the publishing of *Snowfall* in 2012, The New York Times is recognized for significant interactive content, with a taxonomy that differentiates between “multimedia stories,” “data visualizations,” “data-driven stories,” “explanatory graphics,” and “visual and interactive features” (The New York Times 2014). In only a few years, interactive online stories have become de rigueur for large news outlets.

According to *Frontline*’s Raney Aronson-Rath,

> We’re thinking about how to tell stories from the outset not in terms of video vs. audio vs. text, or short-form vs. long-form — but in terms of whatever the right form may be. Maybe a piece of reporting is a visceral, visually compelling story that’s meant to be a documentary. Maybe it will unfold with the most impact as a really thoughtful written piece. Maybe it’s an interactive combination of both. I’m seeing a new openness to this sort of right-form experimentation among journalists (2014).

This “experimentation” includes pushing the boundaries of data storytelling through adding narrative layers and contextualizing quantitative data with qualitative content such as photos, videos, or voice-over narration. Seeking the “right-form” with data invites endless opportunities for innovation. With increasingly powerful browser-based technology and loosening of audience expectations around what constitutes an online story (e.g. no longer “video vs. audio vs. text,” as Aronson-Rath suggests), creators have greater freedom to challenge conventions of online storytelling.
1.1 Research questions and methodology

The current surge in interactive data storytelling represents one manifestation of centuries old efforts at translating data—quantitative and qualitative—into stories. While there is no simple way to translate data into compelling narratives, new data collection and online storytelling strategies foster a mode of communication that can reveal complexities, time-based shifts, and arcane patterns with regard to newly available geolocated data. This thesis investigates interactive storytelling as a mode of communicating data and analyzes trends and opportunities for future innovation. Surveying the field and analyzing specific projects lays the foundation for a design intervention that adds a narrative layer to geolocated, citizen-collected data. The primary research questions of this thesis are:

1. What are the most effective strategies for translating data into interactive online stories that are compelling for wide public audiences?

2. What is the ideal design and functionality of a tool that adds a narrative layer to geolocated, citizen-collected data and widens public access to creating interactive data narratives?

3. What are the implications of this mode of visual representation for public engagement with data?
Methodologically, this research involved interviewing project directors and practitioners, including individuals from organizations such as NPR Visuals, ProPublica, and Internews engaged in innovative interactive and data-driven online storytelling. These interviews provided deep insight into how news agencies and interactive producers frame the field and how resources and technological affordances mediate their approaches. Additionally, close analysis of a range of projects provided a foundation for identifying design insights and opportunities. Analyzing each project’s user experience, narrative arc, information architecture, and underlying code libraries informed the design intervention proposed in chapter four. Lastly, design-based experimentation—including creating wireframes, conducting comparative analysis, and crafting mock stories—provided the basis for the design intervention.

1.2 The challenge of defining data

In addressing various strategies for crafting online narratives from data, delimiting and defining data poses a challenge. Data can range from interviews conducted by a journalist in the field to geolocated photos scraped from the web. As elucidated by Lisa Gitelman in her edited volume, *Raw Data is an Oxymoron*, data is inherently socially constructed and imbued with subjectivity:

At first glance data are apparently before the fact: they are the starting point for what we know, who we are, and how we communicate. This shared sense of starting with data often leads to an unnoticed assumption that data are transparent, that information is self-evident, the fundamental stuff of truth itself. If we’re not careful, in other words, our zeal for more and more data can become a faith in their neutrality and autonomy, their objectivity (2013, 2-3).

---

3 Chapter three, “Design patterns and structural strategies” discusses an analysis of twenty-five projects.
Data are captured and presented by humans who act as interpretive intermediaries. In acknowledging the active process of “capturing” data from the world, Johanna Drucker suggests using the term “capta” rather than “data,” particularly in digital humanities and humanistic inquiry, as “capta” derives from the Latin word meaning “taken” while “data” derives from the Latin word meaning “given” (2011). According to Drucker, differences in the etymological roots of the terms data and capta make the distinction between constructivist and realist approaches clear. Capta is "taken" actively while data is assumed to be a "given" able to be recorded and observed. From this distinction, a world of differences arises. Humanistic inquiry acknowledges the situated, partial, and constitutive character of knowledge production, the recognition that knowledge is constructed, taken, not simply given as a natural representation of pre-existing fact (2011).

While this research does not explicitly employ the word “capta,” the conception of data employed in this research acknowledges data as a construct and humans as playing a significant role in capturing and imbuing data with meaning. This thesis employs a broad definition of “data,” including information collected through sensor hardware, mobile data collection apps, and other quantitative and qualitative data collection strategies. The analysis of interactive data storytelling, however, focuses primarily on projects that incorporate quantitative, geolocated or geographic data (many projects under this scope also interweave photos, videos, text and other content).

1.3 Chapter overview

This thesis is divided into four chapters. Chapter two provides historical and theoretical foundations for today’s innovation in interactive data storytelling, beginning with analyzing a historical precedent, the U.S. Statistical Atlas, to illustrate that the underlying impetus to craft
narratives from data is a longstanding phenomenon. Chapter two then discusses the current proliferation of data, focusing on citizen data collection tools such as Ushahidi and the Open Water Project’s Riffle. A significant portion of newly available data is geolocated, presenting a unique opportunity to translate map-based data into interactive narratives. Finally, this chapter addresses multiple levels of public engagement in acquiring and presenting geolocated data and suggests need for widened access to creating interactive data narratives.

Chapter three provides a survey of the interactive data storytelling field to identify emerging patterns and guiding conventions. Twenty-five projects from a range of sources, including news organizations and independent interactive agencies, provide the basis for this survey. Projects range from long-form interactive stories integrating map-based data such as The New York Times’ *The Downside of the Boom*, to animated data visualization such as Kiln’s *Carbon Emissions: Past, Present and Future*. These projects draw from a range of data sources—including open data, citizen-collected data, and field data—and leverage the affordances of data sources through diverse presentation strategies. Three primary structural strategies arose from this field survey: the guided tour, the scrolling story, and the interactive web documentary. The goals of the survey are to identify design patterns and structural strategies for interactive data narrative; evaluate the strengths and limitations of these strategies; and to inform the design intervention proposed in chapter four.

Chapter four builds upon the field survey by proposing a design intervention—CartoStory—for adding a narrative layer to geolocated, citizen-collected data. The field survey illustrates innovative approaches to crafting narratives from data, but also suggests opportunity

---

4 See Appendix for list of interactive stories included in the field survey.
for greater integration of geographic data with other content and for lowering technical barrier of entry to crafting data stories. This chapter outlines the design process and initial prototyping of CartoStory, an open-source tool that allows users to create a short, narrative tour through map-based data. The goal is to enable creation of engaging narrative experiences from map-based data and to lower barrier of entry for creating such stories. The design process included conducting a comparative analysis of interactive map-based storytelling tools, crafting mock story outputs, and creating wireframes and designs. Design decisions are based on two specific user types and user scenarios: citizen data collectors who are collecting data for a specific organization or issue, and journalists who are crafting map-based stories. This chapter presents the design process and also suggests next steps for further development of CartoStory.

Chapter five, the conclusion, addresses the implications of communicating data through interactive narrative, including potential for advocacy efforts and for public engagement with complex datasets. The conclusion also addresses potential areas for further research and opportunities for design innovation (including streaming real time data, integrating physical user engagement, and creating tools that balance customizability and ease-of-use), thereby pointing toward potential shifts in the coming years.
Chapter two

History, theory, and motivation

Individuals now have unprecedented capacity to source data, whether through collecting data in the field or accessing open data repositories. While individuals can collect and share data, translating data into interactive stories often requires technical expertise and financial resources. This chapter addresses this challenge by outlining the theoretical and practical foundations for today’s innovation in data collection and interactive data storytelling. This section begins with a historical precedent, the U.S. Statistical Atlas, and its parallels with a contemporary interactive story, ProPublica’s Losing Ground, to suggest that the underlying impetus to craft narratives from data is an age-old practice. Next, this chapter discusses the current proliferation of data, particularly focusing on citizen data collection projects and roots in earlier citizen science initiatives. This section is followed by a discussion of the geolocated characteristic of newly available data and the emergence of online interactives that translate geographically specific data into narratives, including an example called Faces of Fracking. The final section discusses public engagement with interactive data narrative and particularly translating data into interactive narratives beyond annotated maps and static data visualizations.

---

5 The term “interactives” is used to refer to interactive online stories.
1.1 Historical precedent

While it is tempting to conceptualize today’s interactive data storytelling as a wholly contemporary phenomenon, media innovations do not arise in a “historical vacuum” (Winston, 1990, 73) and precedents for today’s data storytelling are visible in early data collection and presentation processes. Consider momentarily the parallels and similitude between an early precedent—the Statistical Atlas—and a contemporary analog published by ProPublica—Losing Ground.

The Statistical Atlas of the United States, first published in 1874, exemplifies innovative layering of multifarious data and widespread public access to visual representations of data. The goal of the first Statistical Atlas “was to present a graphic portrait of the nation, and it covered a wide range of physical and human topics: geology, minerals, weather; population by ethnic origin, wealth, illiteracy, school attendance and religious affiliation; death rates by age, sex, race and cause, prevalence of blindness, deaf mutism and insanity; and so forth” (Friendly 2008, 34). While data visualization pre-dated the first Statistical Atlas, the Atlas ranks among early examples of widely available and visually digestible geographic, political, and demographic data. Perusing the early Statistical Atlas provides a rare glimpse into early representation of common resources, ranging from choropleth maps of river basins, forestry, and rainfall to time-series charts illustrating agricultural prices. As in most data visualization, the first Statistical Atlas was imbued with subjective presentation of data; one of the purported goals of the Atlas, according to the U.S. Census website’s History of Statistical Atlases, “was to mark the upcoming centennial by illustrating the progress of the United States over the past century...showing the expansion of settlement by plotting the westward encroachment of population density on maps from each
The early Statistical Atlas exemplifies the power of data visualization for promoting particular socio-political aims. While data visualization certainly pre-dated the first Statistical Atlas, the Atlas is among early examples of widely available, visually digestible demographic data. “The publication of the Statistical Atlas in 1874 was a watershed moment for statistics in the United States. It provided a quick and easy way to interpret comparative statistics on a national level” (U.S. Census Bureau 2015).

As a contemporary analog to the early Statistical Atlas, ProPublica’s innovative Losing Ground interactive story also layers geographic data to convey a particular agenda. ProPublica and The Lens collaboratively produced Losing Ground in 2014. The map-based story uses satellite imagery and personal stories to depict shifts in the Louisiana coastline over the last eighty years. “At the heart of the story is the fact that the Louisiana coastline loses land at a rate equivalent to a football field each hour” (Pitt 2014).

The team behind Losing Ground used historical remote sensing data to illustrate changes to the wetlands and coastline of Louisiana. The interactive map layers visual representation of pipelines, levees, and oil and gas exploration in the region. Losing ground is structured as a “high-level view with eight zoomed-in case studies. The scene of land, marsh and water known locally as the Texaco Canals forms one of the most dramatic examples. Starting with data collected from aerial photography in the 1950s and ending with 2012 satellite data, the layered maps show how the canals sliced up the marshlands and the relocated soil stopped sediments replenishing the land” (Pitt 2014). By weaving personal stories into the interactive map experience, the creators of Losing Ground contextualize potentially dry, historical data with the vibrant stories of people currently affected by receding coastlines. Users can toggle between a 1922 and 2014 map to view the dramatic shift in coastline over the last century. The Losing Ground piece is innovative in
Figure 1 (left) is a page from the earliest Statistical Atlas and illustrates crop distribution in the eastern United States. Figure 2 (right), from Losing Ground, illustrates coastal wetlands and distribution of canals in Louisiana. While the Statistical Atlas depicts layers of data in succession (e.g. side by side imagery), Losing Ground overlays various map layers in the web browser. Both visualizations depict shifting geographic conditions through interweaving various datasets.
At first glance, ProPublica’s *Losing Ground* stands in contrast to the early *Statistical Atlas*. *Losing Ground* focuses on a discrete geographic region (Louisiana’s coastal wetlands), employs contemporary remote sensing technology, and relies on web distribution. The two data presentations, however, rest on similar premises. Both layer multifarious data, albeit through the distinct mediums of print publication and online distribution. Both are map-based and rely on geographic datasets. Both interweave distinct geographic data, including natural resources and human-induced developments. Lastly, and significantly, both initiatives employ data presentation to suggest particular, subjective accounts of shifting social and geographic conditions: *Losing Ground* suggests dire need for coastal restoration, and the *Statistical Atlas* suggests a teleological notion of progress and westward expansion. Despite being published 140 years apart, these two examples of data storytelling represent similar strategies for communicating data on common resources to public audiences.

Today’s innovation in interactive data storytelling, exemplified by such pieces as *Losing Ground*, builds upon the foundations of data communication set through early data collection and visualization initiatives. The parallels between historical and contemporary data storytelling reflects an underlying desire to translate data into digestible, narrative formats.

### 2.2 Proliferation of data

The first *Statistical Atlas* drew from U.S. Census data collected through a time and resource-intensive process; the U.S. Census Bureau’s History of Data Collection describes how “census-takers only asked a handful of demographic questions, processing and tabulating questionnaires occurred at a local level, and publications were relatively limited” (U.S. Census
Bureau 2015). In the nearly century and a half since the first *Statistical Atlas*, however, technological advances have contributed to more efficient and widespread data collection efforts. The last several decades have witnessed significant growth in available data and greater capacity for the public to source, interrogate, and present data. According to the *Economist*, “as the capabilities of digital devices soar and prices plummet, sensors and gadgets are digitising lots of information that was previously unavailable. And many more people have access to far more powerful tools” (2010). Individuals have immediate access to a wealth of data on macro and micro scales, ranging from remote sensing data illustrating deforestation to air quality data indicating pollutants. Furthermore, individuals have the option to access data through open repositories or use readily available tools to collect their own data. New technologies afford the public greater capacity to source data, whether collecting data using a mobile phone app, accessing data through a government open data portal, or gathering data using a “Do It Yourself” (DIY) sensor data collection tool.

A recent Tow Center report on “Sensors and Journalism” asserts that, “community participation in the journalism process has blurred the lines between news professionals and the audience...This social and political context is critical to thinking about how journalists work with, and collect data about, their communities” (Pitt 2014, 152). The shifting boundary between “news professionals and the audience” (Pitt 2014, 152) includes greater public agency in collecting data for interactive stories.

### 2.2.1 Data collection tools

Individuals and organizations can choose from a range of data collection tools. Widespread mobile phones and increasing internet access contribute to democratizing data
collection efforts through open-source data collection tools, including both software (e.g. mobile apps) and hardware (e.g. DIY sensor systems). Prior to these enabling technologies, citizen data collection often necessitated pen and paper recordings or access to expensive (and often proprietary) data collection tools.

One of the most lauded examples of citizen data collection is Ushahidi, a crisis mapping initiative with a suite of open-source tools for collecting data and creating interactive online maps. Ushahidi originally emerged from post-election violence in Kenya after the 2007 presidential election (Meier 2013; Giridharadas 2010). The initial deployment of Ushahidi allowed the public to submit SMS messages containing reports of violence; these reports were aggregated and displayed on an open online map (Ushahidi 2015). In the post-election environment, the tool “collected more testimony — which is what ushahidi means in Swahili — with greater rapidity than any reporter or election monitor” (Giridharadas 2010). In the years since early development, organizations have used Ushahidi to collect and visualize data on issues ranging from post-earthquake damage after Haiti’s 2010 earthquake (Giridharadas 2010) to pollution reports after the BP Deepwater Horizon oil spill (Nienabar 2010).
Ushahidi was an early pioneer in crisis mapping; according to Patrick Meier, "crisis-mapping platforms display eyewitness reports submitted via e-mail, text message, and social media. The reports are then plotted on interactive maps, creating a geospatial record of events in real time" (2013). While Ushahidi ranks among the most well known examples of citizen mapping, a range of other initiatives enable citizens to collect, aggregate, and visualize data. For example, Open Data Kit, a data collection and aggregation tool from the University of Washington, allows people to create custom data collection forms for deployment on mobile phones (Open Data Kit 2015).\(^6\) Open Data Kit allows users to create highly customized forms,\(^6\)

\(^6\) Open Data Kit (ODK) includes ODK Build (online form builder), Collect (mobile app for data collection), and Aggregate (aggregation and visualization platform).
including implementing conditional questions, using XML or an online form builder called “ODK Build.”

In addition to mobile data collection apps, low cost sensor hardware allow people to collect data on environmental issues such as air or water quality. The Open Water Project’s Riffle, for example, is a low cost water monitoring tool designed to measure water conductivity, temperature, and depth (Open Water Project 2015). The Riffle measures “some of the most common water quality parameters, using a design that makes it possible for anyone to build, modify, and deploy water quality sensors in their own neighborhood” (Open Water Project 2015). As described on the Open Water Project website, “traditional water monitoring uses expensive, proprietary technology, severely limiting the scope and accessibility of water quality data” (2015). The Riffle opens up water quality monitoring to groups who cannot necessarily access expensive, proprietary technology, including homeowners, community groups, and environmental advocacy organizations.

---

7 The Open Water Project is a collaboration between Public Lab, Center for Civic Media, and other organizations and individuals working on water quality issues.
These low-cost, open-source data collection tools, including mobile tools like Open Data Kit and DIY hardware such as the Riffle, are rooted in earlier citizen science\(^8\) initiatives that sought to engage citizens in data collection and monitoring for scientific research (Busch 2013). Citizen science involves “communities or networks of citizens who act as observers in some domain of science” (Goodchild 2007, 218) and thereby challenges the notion that only experts have the capacity to collect scientific data.\(^9\) While citizen science began largely as a pen and paper enterprise, web-based technologies such as mobile phones, GPS, and databases streamline data collection and aggregation (Busch 2013; Devisch and Veestraeten 2013; Nielsen 2012). Citizen

---

\(^8\) Rick Bonney termed the coin “citizen science” in a grant application at Cornell University in 1995. In recent years, Bonney shifted towards employing the term “PPSR” (public participation in scientific research), and divides PPSR into three primary categories: contributory projects, co-created projects, and collaborative projects (Busch 2013).

\(^9\) While citizen science projects grant varying degrees of agency to citizen data collectors (some critics suggest citizens are treated as passive data collectors), citizen science projects successfully engage citizens in a range of data collection and scientific research efforts.
data collection tools such as Ushahidi and Open Water Project’s Riffle rely on similar technologies and collaborative principles as citizen science initiatives, enabling greater public participation in data collection.

2.2.2 Open data

In addition to collecting data, communities can freely share and access data online through open data repositories. Open data, defined by the Open Knowledge Foundation as “data that can be freely used, re-used and redistributed by anyone - subject only, at most, to the requirement to attribute and sharealike” (Open Data Handbook 2012, 6) reflects underlying values of collaboration and knowledge sharing.¹⁰ Tools like Open Data Kit and Ushahidi employ open data formats, such as XML¹¹ and GeoJSON¹², to foster data sharing and collaboration. The Open Knowledge Foundation describes financial and technical barriers associated with proprietary file formats: “Using proprietary file formats for which the specification is not publicly available can create dependence on third party software or file format license holders. In worst-case scenarios, this can mean that information can only be read using certain software packages, which can be prohibitively expensive, or which may become obsolete” (Open Knowledge Foundation 2012, 18). Wider adoption of open data formats and open data principles over the last several years strengthens public agency in collecting and sharing data. The public can share

¹⁰ The trend toward open data is a recent phenomenon, with many governments announcing open data initiatives over the last several years. In 2013, Obama signed an executive order to make more government data “open and machine readable” (Cohen 2013). Agencies under the U.S. government—such as NASA, NOAA, and the EPA—build websites and web applications specifically for accessing and exploring open data. Non-government entities also prioritize open data; organizations such as the World Bank and the World Resources Institute (WRI) increasingly make data available online and build custom applications for accessing and interrogating data.

¹¹ Extensible Markup Language

¹² Geographic Javascript Object Notation
independently collected data as well as access open datasets; for example, individuals can access remote sensing data through the U.S. Geological Survey’s online repository or access open environmental data through the Environmental Protection Agency’s “DataFinder” website.

2.2.3 Geolocated data

Many of the aforementioned data collection initiatives—such as Open Data Kit and Ushahidi—associate specific geocoordinates with data points. Mobile computing and geolocation technology allow developers to easily integrate geolocation into data collection platforms. Mobile phones collect geolocation through locating nearby cell towers, crowdsourcing wifi network data, or GPS; remote sensing data is inherently geographic; and even open data portals often feature data sorted by city, state, or country.

Newly available, geolocated data presents a unique opportunity to craft interactive online narratives. Sourcing data is a critical first step, but creating and publishing interactive content opens avenues for widespread public engagement with data. This mode of online storytelling represents an opportunity to translate data on issues that are difficult to represent, such as coastal erosion as illustrated in the Losing Ground example, into publicly accessible narratives.

2.3 Interactive narrative based on geolocated data

Simultaneous with proliferation of data collection initiatives, news organizations, interactive agencies, and independent producers are driving innovation in interactive and data-

---

13 Periscopic, for example, is an agency describing itself as “a socially-conscious data visualization firm that helps companies and organizations promote information transparency and public awareness” (Periscopic 2015) and
driven storytelling. According to a recent report from Columbia’s Tow Center for Digital Journalism, *The Art and Science of Data-Driven Journalism*, “The open question in 2014 is not whether data, computers, and algorithms can be used by journalists in the public interest, but rather how, when, where, why, and by whom. Today, journalists can treat all that data as a source, interrogating it for answers as they would a human” (Howard 2014, 4). Visitors to The New York Times homepage are almost as likely to experience interactive multimedia and data-driven stories as traditional, text-based content. Interrogating and visually presenting data in online interactives is nearly a requisite for news organizations adapting to the rapid pace of online innovation.

Advances in browser technologies and responsive design allow wide audiences to load and experience interactive content, whether on a desktop computer or mobile device. According to Edward Segel and Jeffrey Heer in “Narrative Visualization: Telling Stories with Data,”

Static visualizations have long been used to support storytelling, usually in the form of diagrams and charts embedded in a larger body of text. In this format, the text conveys the story, and the image typically provides supporting evidence or related details. An emerging class of visualizations attempts to combine narratives with interactive graphics. Storytellers, especially online journalists, are increasingly integrating complex visualizations into their narratives (2010, 1139).

Al Shaw, News Applications Developer at ProPublica, emphasized a similar sentiment during an interview about ProPublica’s innovative data journalism and interactive storytelling work:

“Before you had these separate teams. You had these graphics teams and design teams that are designing the website of that organization. And now you’re kind of seeing all that merge together describes its process as converting “raw data into visual and interactive experiences that allow people to empathize and understand” (Periscopic 2015).
so teams that are designing stories are working more closely with teams that are designing graphics.”

Like data visualization, data journalism and interactive storytelling are gaining public familiarity. Crafting interactive narratives based on data lies at the intersection of data visualization, data journalism, and interactive storytelling. While this research does not seek to strictly delimit the field of interactive data narrative, the projects examined in this research are often characterized by strong narrative arc; interweaving of distinct content types; combination of qualitative and quantitative data; and some degree of user interaction. While annotated maps and traditional data visualizations require users to independently navigate and derive insights, interactive data stories provide users with a structured narrative experience.

*Faces of Fracking,* an interactive story about the impacts of fracking in California, exemplifies the potential to weave geographic data into an interactive narrative experience. The story, produced by CEL Climate Lab and Grist, incorporates a map, text, and data visualizations to visually depict the environmental impacts of fracking. Scrolling down the page triggers events corresponding with narrative text; for example, the text, “Here are the LA Basin’s 532 oil & gas wells, where these high-intensity techniques have been used 477 times in the last year alone” corresponds with the appearance of map icons representing the well locations. As users scroll further down the page, the map data shifts to illustrate the level of toxins released from various wells. As the text begins to describe the impact of toxins on local water supply, the map depicts

---

14 The broad spectrum of projects is evident in chapter two’s field survey.

15 *Faces of Fracking* is a larger multimedia story; one component of the project is the narrative visualization discussed in this section.
the locations of groundwater aquifers and the high-intensity production wells and wastewater injection wells in proximity to those aquifers.

*Faces of Fracking* demonstrates the potential for weaving geolocated data into narrative through combining text with interactive data visualizations (the site uses D3.js, a Javascript library for interactive visualizations). Rather than only presenting a text-based story or an annotated map, the site weaves together text with pertinent map data to create a strong visual narrative. The structure and design of *Faces of Fracking* represents one of many emerging strategies for crafting narrative from geolocated data, as will be explored in depth in the field survey in chapter three.

### 2.4 Public engagement with interactive data narrative

As illustrated, the last several years have witnessed a convergence of new data collection strategies and innovation in interactive storytelling. Despite these advancements, however, there remains a missing link between public engagement with collecting data and crafting narratives from that data. The public has tremendous capacity to collect geolocated data, with open-source tools and open data repositories allowing people to source previously inaccessible data. Data acquisition, however, represents only one phase of public engagement with data. After sourcing data, how is this data translated into compelling narratives? How can individuals and organizations leverage innovation in interactive online storytelling to craft visual narratives from data?
Many data collection initiatives present data through annotated maps or static data visualizations. Ushahidi, for example, prioritizes crafting annotated maps based on citizen-collected data. Similarly, Open Data Kit's visualization engine includes three default visualization options: pie chart, bar graph, or map. Despite immense innovation in interactive online storytelling, much of this innovation is driven by news organizations and interactive agencies with financial resources and in-house technical expertise. Individuals and organizations engaged in data collection initiatives often lack technical capacity and resources to craft compelling interactive narratives from the data they collect. Groups using data collection tools such as Ushahidi or the Open Water Project's Riffle should be able to create compelling interactive narratives that guide users through data rather then necessitating significant user navigation.

This challenge—crafting compelling narratives from data (and especially citizen-collected data)—lies at the core of this thesis. This thesis focuses on multiple levels of public engagement—sourcing data, crafting narratives, and viewing story outputs—but focuses primarily on crafting narratives from data. In addition to tools and resources for sourcing data (as discussed in the "Proliferation of data" section), the public needs access to tools for crafting visual narratives from data. The design intervention, discussed in detail in chapter four, addresses this challenge.

---

16 "ODK Aggregate" is Open Data Kit's data aggregation tool that allows users to create simple visualizations.
Chapter three

Design patterns and structural strategies

The field of interactive data narrative has grown significantly over the last several years, with sizeable interactive departments at many major news outlets. Interactive data narrative, however, is a nascent field; thus, there are few guiding conventions and best practices for sourcing and presenting data. While best practices of web design and development—such as responsive design, minified source code\(^\text{17}\), and intuitive user navigation—apply to interactive storytelling, producers also regularly introduce new and experimental design strategies. Browser updates and new technical affordances allow project producers and developers to innovate on existing strategies.

This chapter addresses the challenge of identifying emerging patterns and guiding conventions by surveying the field of interactive data narrative. The aim of this analysis is to inform the design and functionality of the tool, proposed in chapter four, for adding narrative to geolocated, citizen-collected data.

The first section, “Overview of projects,” describes the projects included in the survey and the basis for selecting these projects. The “Structural approaches” section describes three structures for interactive data narrative: the guided tour, the scrolling story, and the interactive web documentary. Next, “Data sources” provides an overview of data used in projects—including open data, citizen-collected data, and field data—and how projects leverage the affordances of

\(^{17}\) “Minified” refers to code where extra characters are removed to optimize load time.
various data sources. The final section, “Insights and opportunities based on field survey,” highlights areas of opportunity including highly guided narrative experiences and lowered technical barrier to entry. The underlying goals of this survey are to identify design patterns and structural strategies for interactive data storytelling; ascertain the efficacy of these techniques; and ultimately inform the design intervention proposed in chapter four.

3.1 Overview of projects

Surveying the field involved selecting twenty-five interactive storytelling projects that integrate data through a range of strategies. Projects were selected from news outlets (The New York Times, The Guardian, NPR, National Geographic, and the Texas Tribune), interactive agencies (Periscopic, Kiln, and Zemi Media), and public media (NFB, Localore). A small segment of projects were created by independent producers or by groups at the Datalore Hackathon\(^\text{18}\) at the MIT Media Lab. Analyzing these projects included conducting interviews with producers and developers behind a selection of the projects; analyzing the design, structure, and narrative strategy of each project; and running projects through site analysis tools such as Google’s Page Speed Insights.

Each project includes a geographic component but employs divergent narrative and structural strategies. Selecting projects featuring geographic data proved simple; as discussed in chapter two, the rise in distributed data collection and mobile and ubiquitous computing allows for highly geographically specific data. Citizen-collected data is frequently associated with

\(^{18}\)Datalore Hack was a three-day data storytelling hackathon I co-organized at the MIT Media Lab in January 2015. The hackathon was a collaboration between the MIT Open Doc Lab, MIT Center for Civic Media, Harvard Bok Center, Non-Fiction Cartel, and Storycode Boston.
specific geo-coordinates; remote sensing data is inherently geographic; institutional datasets often feature city, state, or country level parameters. Thus, a significant segment of data-driven interactive stories are geographic or explicitly map-based.

The projects surveyed employed highly varying storytelling strategies. Some projects weave data visualizations into text, video, and photo content. For example, Return to Elwha is an interactive documentary focusing on the removal of the Elwha River dam—the “world’s largest dam removal” (Nijhuis 2014)—featuring video, text, and animated maps in the form of a long-form scrolling story. Other projects are highly guided, narrative experiences featuring voiceover. For example, The Guardian produced an interactive in collaboration with the World Resources Institute called Carbon Emissions: Past, Present and Future. The interactive site illustrates shifting CO2 output on a country-by-country basis over the last 150 years and allows users to watch an animated tour or independently navigate the visualization.

The projects chosen are not a representative sample, but are a curated set of projects chosen to reflect diverse presentation strategies and data sources. This survey explicitly includes projects with a wide range of storytelling strategies, underlying architecture, and data sources in order to broadly frame the field.

3.2 Structural strategies

While the projects surveyed employed a range of user experience and design techniques, three primary structural approaches emerged from the survey: the guided tour, the scrolling story, and the interactive web documentary.
3.2.1 Guided tour

The “guided tour” shepherds users through a highly guided experience through data-driven narrative. With guided tours, users frequently have the option to click a “play” icon to watch a visualization unfold and are subsequently presented with the option to navigate independently. This structure caters to distinct user types with varied engagement levels, as users can passively experience the story or navigate in more depth. In some cases, the guided portion of the experience consists of a short introduction with voiceover narration. In other projects, the guided experience lasts several minutes and guides users through an in-depth exploration of a visualization.

The interactive agency Kiln frequently employs the guided tour technique.\(^{19}\) *Carbon Emissions: Past, Present and Future*, developed by Kiln with the World Resources Institute, allows users to watch an animated tour of the data visualization and then independently explore the visualization in more depth. This structure is similar to playing video content, but offers users the option to pause the animation at any moment and independently navigate.

This visualization focuses on shifting carbon dioxide emissions among several countries over the last 150 years. Voice-over narration describes historical emissions trends and contextualizes data seen on-screen. The guided tour describes increase in emissions starting with the industrial revolution; the emergence of the U.S. as the top emitter at the turn of the 20th century; and the emergence of China and other countries as recent top emitters. Voiceover narration contextualizes the changing emissions levels of various countries, situating shifting emissions within broader global socio-economic dynamics. The animation and voice-over adds a

---

\(^{19}\) Kiln is an agency that frequently produces interactive content for The Guardian. According to their website, “Kiln produces interactives, maps and visualizations that bring complex data and stories to life” (Kiln 2015).
narrative arc to the data visualization, elevating it from a visualization to an interactive narrative experience.

Kiln frequently employs the guided tour technique and has developed and released open-source tools for others to use similar techniques. Kiln, for example, released Talkie, a Javascript library for adding voiceover narration and animated effects to visualizations. Kiln describes Talkie as “a tool for producing pieces of content – ‘Talkies’ – that combine the accessibility and story-telling strength of a video with the depth and explorability of a data-rich interactive visualisation. At its simplest, creating a Talkie means adding an audio voiceover to an interactive map or visualisation so that users can be ‘talked through’ the content rather than approaching it cold and doing all the interpretation themselves” (Kiln 2015). Tools like Talkie allow others to create interactives with a strong narrative arc and highly guided user experience. In releasing Talkie open-source, Kiln hopes “that releasing an easy-to-use library will help the Talkie concept become adopted more widely. Almost every interesting data interactive would be improved by the addition of a short Talkie introduction” (Kiln 2015).

Guided tours strengthen the narrative arc of interactive content and also widen accessibility to engaging with interactive data narratives. Data visualizations frequently demand a high degree of data literacy and can be confusing for users unfamiliar with interpreting data. Guided tours orient users to visualizations and provide interpretive context, although they also potentially discourage users more interested in a self-guided interactive experience.

3.2.2 Scrolling story

The second strategy that emerged from the field survey is the scrolling story. These stories are organized in a linear, top-to-bottom structure and often allow user navigation on specific
page elements. Usually the only required action to experience these stories, however, is scrolling. *Snowfall*, published by The New York Times in 2012, was among the first projects to familiarize audiences with the scrolling, long-form multimedia narrative format. According to Nieman Lab, “It wasn’t long ago that news companies were hesitant to put significant content ‘below the fold’ - the old newspaper metaphor moved from newsprint to screenfuls. The BuzzFeeds and Snow Falls of the world have taught publishers to think of scrolling less as a hindrance and more as a useful, tactile part of the content consumption experience” (Benton 2013).

Scrolling is an increasingly common format across the web, from Bootstrap websites to news apps with continuously loading content on scroll. After the success of Snowfall, a start-up based in New York created Scrollkit, a tool for creating Snowfall-like stories through a drag and drop interface. An article in TechCrunch recounted how the founder of Skrollkit “recreated the Snow Fall piece using Scroll Kit to show that you didn’t need an army of developers or designers to create the same type of interactive storytelling. In fact, the tools exist today to build other compelling narratives that take advantage of the combination of text, and video, and images” (Lawler 2013). Scrollkit’s popularity illustrates the growth of the scrolling format and the potential to create tools that widen access to interactive storytelling, which will be discussed in chapter four.

ProPublica’s *Louisiana’s Moonshot* is an example of a scrolling story incorporating map-based data, although it maintains map data and text as largely discrete entities. ProPublica, an “independent, non-profit newsroom that produces investigative journalism in the public interest” (ProPublica 2015), produces data-driven content including data visualizations and news applications exploring public interest topics as varied as health data breaches and species extinction. ProPublica produced *Louisiana’s Moonshot* as the follow-up to the *Losing Ground*
story discussed in chapter one; the story focuses on a fifty billion dollar plan to restore Louisiana’s coastal wetlands. As depicted in Losing Ground, Louisiana’s wetlands are at risk of disappearing entirely. The state of Louisiana recognizes the potential environmental and economic disaster and has initiated a “moonshot” plan to restore the state’s coastal wetlands. Louisiana’s Moonshot, created by Al Shaw and Brian Jacobs from ProPublica and Bob Marshall from The Lens, tells the story of this moonshot plan through weaving text, photos, and interactive map data into a long-form scrolling story.

*Louisiana’s Moonshot* focuses on three primary restoration projects: a marsh creation project, a diversion project, and a river modification project. In stark contrast with Losing Ground, in which the interactive map remains visually central to the user experience, Louisiana’s Moonshot allows users to scroll through the content of the story. As users scroll down, they have opportunities to navigate interactive maps. For example, after the initial introductory text, users can navigate an interactive map featuring four primary layers: “today’s wetlands,” “critical projects,” “one possible future,” and “explore the plan.” At first glance—without navigating through the map layers—users can ascertain the significant expanse of wetlands around coastal Louisiana and the importance of wetlands for buffering Louisiana from storm surges (an annotation on the default layer points out the importance of this buffer). If users choose to navigate the map, they can explore current and future restoration plans, view a possible outcome for 2060, or explore the minutiae of different elements of the restoration plans (this map layer is the most detailed, with dozens of annotated hovers). Throughout the story, there are a total of

---

20 Integrally tied to healthy fisheries and sustainable natural resources, receding wetlands have an immense environmental and socio-economic impact. For example, Louisiana’s shrimp fishery (the last thriving shrimp fishery in the U.S.) relies on healthy tidal zones made possible by wetlands.

36
four interactive maps illustrating the primary restoration projects as well as possible future outcomes.

Fig. 3.1: The first interactive map in *Louisiana's Moonshot*

The interactive maps woven into the story allow users to quickly glance and continue down the story, or pause and explore the data in more depth. This storytelling structure caters to different user engagement levels and user preferences; users can experience the complete story without a single click, or they can spend significant time exploring data depicting implications of the restoration projects. Furthermore, while the interactive maps contain significant data, minimal clicks are required to fully explore the maps. The first interactive map (described above)
allows users to click on the four map layers but otherwise requires no clicking; users can explore more detailed data through hovering on map elements.

A lasting tenet of effective information design, first proposed in 1996 and more recently popularized by information design expert Alberto Cairo, is “overview first, zoom and filter, then details on demand” (Shneiderman 1996, 337). While the Moonshot interface does not allow users to zoom, the interactive maps provide an “overview” and then allow users to access “details on demand.” By simply glancing at the default map layer, users gain a high-level overview. Users with higher interest and commitment can access a wealth of details on each of the three primary restoration plans, including hundreds of annotated hovers. One of the successes of this project lies in the presentation of high-level map data that provides an entry into detailed data on each of the three primary restoration projects.

While the map data and text are largely separate in the Louisiana’s Moonshot scrolling story, the interface features one strategy for integrating narrative text with map data: particular words are highlighted and contain a map icon, and when users hover on these words a visual pointer highlights particular map elements. Thus, while reading the narrative text users can hover over words for enhanced contextual understanding. This interaction design technique links textual content with map data. Louisiana’s Moonshot demonstrates how interactive producers can add narrative to map-based data through weaving text, photos, and map data into long-form scrolling stories.

Another example of a scrolling story structure is The New York Times’ The Downside of the Boom, about fracking operations in North Dakota and associated spills and environmental pollutants. The story integrates texts, photos, and data visualizations in a long-form scrolling format. Users can scroll through the story and have the option to intermittently watch animated
data visualizations. For example, the story includes an animated visualization illustrating location of spills from 2006 to 2014 throughout North Dakota. Another animated visualization illustrates distribution of oil wells and associated lines that “connect the fracking operations to surface wells” (Sontag and Gebeloff 2014).

As illustrated in *Louisiana’s Moonshot* and *The Downside of the Boom*, news organizations frequently employ the scrolling story format. Scrolling stories often employ Javascript libraries for event triggers and parallax effects. For example, Skrollr.js is frequently used for parallax scrolling and changing CSS elements of specific page elements, thereby triggering animation effects upon scrolling. One of the strongest aspects of the scrolling story is the intuitive user experience. Users have one primary mode of navigation (downward scroll) and are unlikely to face confusing choices or navigation challenges. Scrolling stories demand minimal interaction from users and have a linear path, but take advantage of the affordances of the web through multimedia content and event triggers at particular vertical scroll positions.

### 3.2.3 Interactive web documentary

The third primary structural strategy is the interactive web documentary, which generally translates data and other content into a full-browser, immersive web format. This format is less clearly defined than the guided tour and the scrolling story, but interactive web documentaries frequently present content on standalone websites rather than being hosted on a conventional news site, and integrate diverse content types such as video, photos, text, and data visualizations.

An exceptional example of interactive web documentary incorporating data is NPR Visuals’ *Planet Money Makes a T-Shirt*. In 2013, NPR’s Planet Money set out to tell the story of how a cotton t-shirt is produced. Inspired by Pietra Rivoli’s book, *The Travels of a T-Shirt in the*
Global Economy, Planet Money tracked the global processes behind the production of a t-shirt, ranging from cotton production to shipping containers to Bangladesh factories. The Planet Money team ran a kickstarter to fund the story and raised a remarkable $590,000, more than half a million dollars above their $50,000 goal (Kirkland 2013). The project began with Planet Money crafting a series of podcasts around the making of a t-shirt, from production of cotton to distribution. To produce the web component of the project, NPR's Planet Money team collaborated with NPR Visuals, who drove much of the interaction design and visual narrative. As a result of the public support and financial resources allocated to the project due to the successful kickstarter, the NPR team was able to weave the content into a web documentary rather than presenting content primarily as text or video.

Planet Money Makes a T-Shirt is not driven primarily by data, but incorporates data visualizations and infographics into the interactive experience. The NPR team devoted significant time and resources to creating the web documentary, spending months undertaking a rigorous user-centered design process. Brian Boyer, Director of NPR Visuals, describes the initial phase of projects as heavily “user-centered:”

What we do is we start every project with a user-centered design exercise. It's changed over time but the general idea is that we want to reset people's expectations a little bit. Frequently when we have a fun idea for a piece it's sort of like, ‘Hey, I've got these five stories, let's put them on a map.’ Then you can click on the map and you can read the thing about the story or something like that. That's a very common first idea...Our user-centered design process helps you reset some expectations.

For Planet Money Makes a T-Shirt, the team steered away from common “first ideas.” The result is a five-chapter web documentary following several phases in the making of a t-shirt, from growing and processing cotton to shipping t-shirts around the world.
The documentary centers on short video vignettes that provide an entry into each chapter. Following each video is a section with narrative text interspersed with photos, data visualizations, and infographics to provide visual illustrations. For example, the chapter titled, "People" includes a data visualization illustrating monthly minimum wage in various countries. This visualization illustrates that Bangladesh, home to a large garment industry, ranks lowest with a $68 monthly minimum wage. Each visualization contains clearly labeled data sources; the minimum wage data, for example, came from the International Labor Organization. While data visualizations are a secondary component of the documentary, Planet Money Makes a T-Shirt illustrates the potential for user friendly, intuitive interactive web documentary weaving together video, text, and data.

Fig. 3.2: The intro to the “People” chapter in Planet Money Makes a T-Shirt
**Planet Money Makes a T-Shirt** is a global story, and much of the data and content presented is geographic. The documentary presents data in various formats (including bar charts, line charts, and annotated photos) but does not feature any maps. Maps are often the default presentation format for geographically specific data, but not always the optimal choice. In *Planet Money Makes a T-Shirt*, the choice of visualization formats matches the message conveyed. For example, the bar chart illustrating global monthly minimum wage visually emphasizes the low minimum wage in Bangladesh, as Bangladesh is hierarchically the last country on the bar chart. Furthermore, the geographic context for the story arises in the video vignettes, which are titled with their location. By focusing on telling human stories through strong narrative, and very
selectively using visualizations and infographics to convey geographic patterns, the team crafted an emotionally engaging web documentary.  

Interactive web documentaries incorporating data often integrate subtle interaction design techniques for guiding users through content. One technique employed in Planet Money Makes a T-Shirt involved triggering a page scroll upon finishing a video, so user attention is directed further down the page without user action. The team used Smoothscroll.js (a Jquery plugin for smoothly scrolling down a page) and Froogaloop.js (a Javascript library that works with Vimeo’s API) to automatically shift users down a page upon completion of a video.

Creating interactive web documentaries incorporating data often involves significant labor and resources. Only over the past few years have news organizations opted to produce interactive web documentaries, and it remains a rarity. Planet Money Makes a T-Shirt is an exception to the norm for NPR and for news agencies more generally; although its success portends potential growth of web documentary in news production, the immense resources necessary to create web documentaries precludes most individuals and organizations from producing such content. A thorough design process and custom development are at the heart of Planet Money Makes a T-Shirt, and the story could not be replicated purely with out-of-the-box tools.

---

21 According to Brian Boyer, Planet Money Makes a T-Shirt was a “turning point” for the NPR Visuals team. The NPR Visuals team was actually formed during the creation of Planet Money Makes a T-Shirt. Brian Boyer explained, “there was a team called multimedia that was photo and video. There was a team called news apps, the team I founded, that looks at graphics and data, interactive. In December 2013, in the middle of the Planet Money project...we merged the two teams together and created a single team called Visuals that is photo, video, graphics, data, interactive.”
3.3 Data sources

The projects surveyed include a range of data sources; some projects draw from a single data source, while others layer multiple datasets. Some of the projects reflect rising enthusiasm around new approaches to sourcing data for stories, whether crowdsourcing data from audience members or gathering data through DIY data collection tools in the field.

Several of the projects surveyed illustrate the willingness of journalists to employ novel approaches to acquiring data. For example, to create *Louisiana’s Moonshot*, the team travelled to Louisiana to gather aerial imagery (remote sensing data) through kite and balloon mapping. The ProPublica team needed high quality, contemporary aerial imagery to examine the efficacy of restoration efforts happening over the last few years (Shaw and Jacobs 2014). While they could obtain satellite imagery from NASA’s Landsat (which they used for *Losing Ground*), this satellite imagery would not provide sufficient detail. Alternately, they could draw from imagery available from USGS Orthophotography initiative, but this imagery was about two years old and would not reflect the latest phases of restoration efforts. According to the team, “What we needed were high-resolution aerial photographs of these areas right now. In remote sensing terms, what we needed was the holy grail: high temporal, spatial and spectral resolution. Like the old adage about fast, cheap and good, you don’t often get all three” (Shaw and Jacobs 2014). To solve this problem, they teamed up with Public Lab, an environmental monitoring non-profit organization that, among other initiatives, gathers aerial imagery through a process called “kite mapping.”

Public Lab “develops and applies open-source tools to environmental exploration and investigation. By democratizing inexpensive and accessible Do-It-Yourself techniques, Public Lab creates a collaborative network of practitioners who actively re-imagine the human
relationship with the environment” (Public Lab 2015). Public Lab takes an innovative approach to developing DIY kits and widening access to environmental data collection.

ProPublica team members travelled to Louisiana and, with the help of Public Lab, flew kites and balloons equipped with cameras to gather images, including both visible and infrared light (infrared is necessary to discern land and water bodies). The team faced some challenges acquiring high quality imagery via kite mapping; in an interview, ProPublica News Applications Developer Al Shaw said, “We had varying success...the first day we went out to shoot Lake Hermitage we didn’t actually have too much success because we had to use a kite that day because it was too windy to use a balloon. And the kite, we couldn’t get that much altitude with it, so it didn’t work that well. The second day...that was a lot better. We could actually get to the same altitude that you see in a NAIP\textsuperscript{22} image.” The team augmented the imagery they gathered with orthorectified imagery from USGS and imagery from DigitalGlobe, a company that provides high-resolution satellite and aerial imagery. Shaw also pointed out the significant labor involved in processing imagery: “Both of them required a lot of post-processing. We had to georeference the images ourselves and warp them because sometimes when the balloon moves around at oblique angles, so it’s not always shooting down at nadir. So sometimes you have to skew it a little in Photoshop to get it to line up.”

ProPublica’s approach to gathering visual data for the Moonshot story is one example of a larger underlying shift toward journalists and citizens acquiring data themselves through newly available tools and data collection techniques. With the rise in citizen data collection tools and open-source technologies that foster decentralized data collection, such as balloon mapping,

\textsuperscript{22} NAIP is the acronym for National Agriculture Imagery Program, an organization that regularly acquires aerial imagery.
individuals and organizations can source data independently rather than relying on institutionally sourced data. Independently collecting data, however, involves the challenge of ensuring accuracy and validity of data; ProPublica faced this challenge when acquiring aerial imagery via kite and balloon mapping, and turned to other aerial imagery repositories to augment their independently sourced data.

Despite the challenges of accuracy, ProPublica’s willingness to travel to the field and source data is indicative of news agencies increasing use of new tools and techniques for obtaining data. The projects analyzed in the field survey draw from a range of sources, and many, like the ProPublica example, use creative techniques for acquiring data.

3.3.1 Open data

Of the projects surveyed, a large segment draw from open data. Open datasets are an accessible option for interactive producers because minimal resources are necessary to download open data (especially when compared with collecting data in the field or mobilizing citizen data collectors).

Many of the interactive stories surveyed use data from government agencies and other institutions. For example, Carbon Emissions: Past, Present and Future, the interactive story illustrating shifting C02 emissions discussed in the “guided tour” section, draws on a range of open data including data from the World Resource Institute’s CAIT (Climate Analysis Indicators Tool), the IPCC (Intergovernmental Panel on Climate Change), and the CDIAC (Carbon Dioxide Information Analysis Center). A data-driven story by The New York Times called Climate Change Threatens to Disrupt the Range of Birds uses data from the National Audubon Society to illustrate shifting bird ranges due to climate change and other factors (the story
illustrates shifting bird habitat by juxtaposing bird range in North America in 2000 and projected range in 2080).

Stories drawing from open data, especially those produced by news outlets, often contain explicit citation of data sources and links to original data. In the case of Kiln’s *Carbon Emissions: Past, Present and Future* story, a “sources and info” link leads to several paragraphs describing and contextualizing the data sources. For example, the CDIAC data is contextualized as follows: “Emissions from deforestation and other land use change are not shown in the country bubbles but are included at the global level in the cumulative total in the budget view. Data source: CDIAC, Carbon Flux to the Atmosphere from Land-Use Changes 1850–2005. The visualisation assumes global land use emissions remain unchanged between 2005 and 2011” (World Resources Institute 2015).

Clearly explaining visual treatment of data (e.g. “emissions from deforestation and other land use change are not shown in the country bubbles but are included at the global level”) and stating assumptions (e.g. “global land use emissions remain unchanged between 2005 and 2011”) insures transparency and reduces potential misinterpretation. The reduced potential misinterpretation is two-fold. First, by taking steps to publicly contextualize data sources, interactive producers face greater public accountability and may reflect on their representational choices. Second, visitors to the interactive story can access detailed information about data sources if they face confusion or questions about the visual representation of data. The New York Times story on bird range, however, only cites the data source, The National Audubon Society, without providing a link or contextualization for the dataset (the piece does specify that the data reflects range during summer months only). While there are no strictly established standards for citing and contextualizing data sources, providing clear explanations of data sources and data
manipulation are critical for data transparency. Linking to original data sources is particularly relevant for open, publicly available datasets.

The movement toward open, machine-readable data—whether from governments, non-profit organizations, or other entities—opens avenues for crafting stories from data without the overhead involved with purchasing proprietary data or independently collecting data. Of the projects analyzed in the field survey, the majority either relied primarily on an open data set or used open data to augment other data sources.

### 3.3.2 Citizen-collected data

Of the projects surveyed, only five projects incorporate significant citizen-collected or citizen contributed data. Despite the affordances of web and mobile technologies, mobilizing citizens to collect data often requires significant energy and resources.

Of the projects evaluated, Localore’s *iSeeChange* is among the best examples of effective citizen data collection. Localore, an initiative of AIR Media Works, upends the traditional public media model by matching independent producers with public media outlets around the U.S. to create innovative, interactive projects (Localore 2015). *iSeeChange*, led by producer Julia Kumari Drapkin, is a flagship LocaLore project that crowdsources citizen contributed climate change commentary onto an interactive website modeled after an almanac. Described as “a groundbreaking environmental reporting project that combines citizen science, participatory public media, and cutting-edge satellite and sensor monitoring of environmental conditions” (*iSeeChange* 2015) the site allows citizens to contribute short comments to the site with attached photos, and also comment on others’ posts.
Drapkin produced the piece while living in Colorado for a year working with the local public radio station in Paonia, a town of 2,000 people. The first iteration of *iSeeChange* involved citizens contributing content via mobile phone, but Drapkin quickly realized that citizens were not motivated to contribute content via mobile device. While citizens of Paonia would chat on the street or submit comments to Facebook, they were not likely to contribute climate related observations using a mobile phone. As a result, Drapkin shifted her model so citizens could contribute stories via Facebook, the iSeeChange website, and by calling the public radio station (Peach 2012; Osenga 2013).

The challenge of motivating participation that Drapkin faced is common to citizen data collection efforts. *Why, and how*, should citizens engage in data collection or contribute content? Answering these questions requires understanding the community contributing content and designing collection and contribution systems in accordance with that particular community’s needs. Drapkin iterated upon the original *iSeeChange* design to match community needs after realizing that mobile contributions were not a viable option. Citizen data collection projects require significant levels of involvement (e.g. spending a year in a small community like Drapkin) and direct engagement, which is not always possible for projects with budget and time constraints.

Another model for citizen data collection involves gathering sensor data. The potential of DIY hardware and crowdsourced sensor data are often touted as the future of data journalism; according to a recent report on sensors and journalism from the Tow Center for Digital Journalism, “We are in an era in which reporters are hungry for data, and increasingly expert in using it; in an age when sensing technology is developing radically and permeating every aspect of modern life. Those trends, taken together in journalism, have produced new demands that
sensors might meet, opportunities that might be exploited, and benefits that might be realized” (Pitt 2014, 21).

Projects that incorporate citizen-contributed sensor data, however, often require significant time, resources, and technical expertise. WNYC’s Radio Lab Cicada project ranked among the first and most successful examples of a data journalism project incorporating citizen-collected sensor data. The WNYC team publicized instructions for building a simple temperature sensor for use in outdoor soil; participants could build their own sensor circuit and contribute their data to WNYC’s online map. Through aggregating soil temperature, the project tracked emergence of Cicadas in 2013 (an event that occurs every 17 years). John Keefe, the project’s director, conceived of the idea at a hackathon: “At a public radio hackathon, Keefe suggested a project in which WNYC members would buy temperature sensor hardware and report their findings to the station in order to predict the fateful day” (O’Donovan 2013). On the Cicada Tracker site, users can view a map of crowdsourced readings and can click on each point to view either temperature readings or cicada sightings. Visitors to the site can also access instructions for building their own Arduino-powered temperature sensor.

The Cicada project succeeded in engaging thousands of public radio listeners in hardware hacking and citizen science. The success may be partially due to the popularity of the show Radio Lab, one of the most widely listened to via podcast, but there is a clear underlying interest in incorporating sensor data into data-driven stories. The project actually incorporates multiple avenues for citizen participation: building sensors and contributing sensor data, attending a hackathon and engaging in hardware development, and contributing cicada sightings through an online form on the project website. Given the challenge of motivating citizen engagement, part of the Cicada project’s success may be due to the various avenues for citizen engagement. Rather
than offering the public just one avenue for engagement, people could engage by simply visiting the website and contributing a sighting, or could engage more deeply by building their own sensor and engaging in a wider citizen science community.

While both *iSeeChange* and *Cicada Tracker* effectively aggregate citizen-collected data, neither presents data online through a highly narrative experience. Unlike many of projects evaluated, these sites provide repository of stories and citizen-collected data rather than a cohesive narrative experience. Weaving citizen-collected data into interactive narrative experiences is one of the greatest challenges in the realm of data-driven interactive storytelling, and one that will be addressed in greater depth in chapter four.

### 3.3.3 Field data

Field data is an umbrella term for data collected by individuals or organizations “in the field.” This could include collecting audio and video material or other forms of qualitative and quantitative data. Unlike proprietary data or open data collected through online research, field data involves capturing data through fieldwork rather than office-based work.

Among the projects evaluated, a significant portion incorporate field data. For example, The Guardian’s *America’s First Climate Refugees* weaves together text, maps, data visualizations, and audio-visual content into a long-form scrolling story. The production team travelled to Newtok, Alaska to gather field data, including interviews, photos, and video. This story relies on personal stories of adaptation to climate change, so gathering field data provided the narrative structure and key voices. The story incorporates data visualizations drawing from available datasets, such as climate data from weather.com and forecast.io, but the personal voices of Newtok community members drive the story.
There are important distinctions between field data and citizen-collected data. First, field data is usually sourced by professionals (e.g. professional videographers collecting video interviews) while citizen-collected data is sourced by amateurs (e.g. citizens collecting temperature readings through DIY circuits). Another distinction relates to scale of data collection; if numerous individuals collect data using a similar collection protocol—such as deploying mobile data collection forms or collecting sensor data—this data is more aptly described as citizen-collected data.

Field data is not restricted to qualitative data such as video and audio interviews. Field data could include quantitative data, such as survey data or environmental sensing data. For example, the creators of Texas Tribune’s *Disappearing Rio Grande* included water quality data in their depiction of shifting environmental conditions on the Rio Grande. To produce the project, a reporter and photojournalist at the Texas Tribune traveled down the Rio Grande “by foot and small boat,” collecting stories, photos, and environmental data along the way (The Texas Tribune 2015). In the final interactive story, water quality measurements—including conductivity, pH level, dissolved oxygen, and turbidity—are featured alongside photos and text content. For users interested in more detailed explanation of water quality metrics, a link labeled “What do these numbers mean” provides context for each water quality parameter. Ultimately, the story is driven by both qualitative and quantitative field data.

Field data ranks among the most common data sources featured in interactive online stories, particularly among interactive web documentaries and long-form scrolling stories, and is often combined with other data sources.
3.5 Insights and opportunities based on field survey

As described, the field survey illuminated several structural techniques for creating interactive data narratives, as well as unique approaches to sourcing data. The field survey suggested several other areas of innovation and opportunity, including greater integration of geolocated data with other content; highly guided user experiences; open development processes; responsive story output; and lower technical and financial barrier to entry. These areas of opportunity broadly inform the design intervention proposed in chapter four.

3.5.1 Greater integration of geolocated data with other content

The projects surveyed reveal the potential for increased integration of geolocated data and narrative to compel audiences to more deeply engage with data. Many of the projects, including Louisiana’s Moonshot and Planet Money Makes a T-Shirt, maintain map data and data visualizations as discrete on-screen entities largely separated from text, photos, and other content. There is potential for greater integration of data with text and other content. A small segment of projects analyzed deeply integrate geolocated data into the interactive narrative experience. Faces of Fracking ranks amongst the best examples of integrating map data with text and other content; as users scroll down the story, map data and data visualizations dynamically react to the scrolling action. There is significant potential for projects that push the integration of geolocated data with text and other content, thereby presenting a cohesive narrative experience that facilitates user engagement with all content rather than focusing on a single content type.
3.5.2 Highly guided user experience

The field survey also surfaced the strength of highly guided narrative experiences. Stories surveyed range from those with low user agency (minimal links and a highly guided user path) to high user agency (many links and possible user paths). Although projects that provide highly guided narrative experiences may seem antithetical to the premise of “interactive” storytelling, these projects often provide engaging and intuitive user experience. According to Brian Boyer, who directs NPR Visuals, “frequently interaction gets in the way of the story instead of helping it.” Boyer’s sentiment points toward the power of highly guided narrative experiences as compared with narratives that necessitate significant user interaction and decision-making. NPR Visual’s Planet Money Makes a T-shirt web documentary, for example, limits user options and provides a linear path: “Despite the wealth of features, the narrative remains linear: At no point does the reader face tough choices about whether to continue reading or to pause for a video in the middle of the text...the interactive documentary purposefully limits the number of anxiety-inducing choices while still giving readers control over the pacing” (Kirkland 2013).

According to ProPublica’s Github guide for “The Design and Structure of a News Application,” creators of news apps should “allow free-form exploration but use element placement and size to make the pathway that tells your story the most prominent and easiest one.

---

23 This spectrum of user agency draws partly from NPR Visuals user-centered design process in creating the Planet Money Makes a T-shirt interactive documentary. Wes Lindamood, interaction designer on the project, describes this spectrum in a blog post about the making of Planet Money Makes a T-Shirt; as part of the design process, the NPR Visuals team examined existing projects and placed them on a spectrum from “user directed” to “narrative journey” (Lindamood 2014).

24 News apps are data-driven stories; According to ProPublica, “News apps tell stories. They’ve got much of the same structure as any news story. They’ve got the graphical equivalent of ledes and nut grafis. At their best, they help a reader to find their personal stories in a large data set and to understand the story you’ve reported using the example of themselves and their own community. A great news application lets a reader understand new concepts by relating them to their own experiences” (Klein 2013).
Remove extraneous elements that confuse the reader or send them on wild goose chases” (Klein 2013). The guided tour and scrolling story structures both offer highly guided narrative experiences and provide clear “pathways” for users. Projects that fall on both ends of the spectrum—offering users the option to passively experience a highly guided narrative or independently navigate—cater to different user types and commitment levels. Some interactives, such as Kiln’s “guided tour” animated data visualizations fall on both ends of the spectrum, allowing users to experience the entire story without making choices while simultaneously offer users the opportunity to explore on their own. Stories that simultaneously provide users with a guided experience and the option to self-navigate are challenging to design but also enticing to a range of user types.

3.5.3 Open development processes

Many of the projects surveyed come from news organizations and interactive producers operating according to open development principles. News organizations—including NPR and ProPublica—are embedded in an ecosystem of innovation that continually iterates upon existing development processes by creating tools, code libraries, and templates. This ecosystem of innovation is largely open-source. ProPublica, for example, prioritizes open innovation, with a repository of tools and “Nerd Guides” publicly available via GitHub. In addition to publishing source code for many projects, ProPublica releases open-source tools, resources, and documentation. Al Shaw explained that, “Open source to us is not just literally releasing code. Open source is writing documentation, it’s being active in the issues, giving active responses to
people who are filing bugs, and encouraging people to participate in the process. It’s not just literally having code on the internet.”

NPR Visuals also publishes many of their tools and projects on Github. The entire source code for the *Planet Money Makes a T-Shirt* web documentary, for example, is available via Github. Although the *Planet Money Makes a T-Shirt* Github repository contains the stipulation that the code cannot be re-used without permission from NPR, the repository nonetheless provides the public with an immense amount of insight into the technical implementation of the project, including documentation and specific code libraries used (e.g. froogaloop.js, fitvid.js, and smoothscroll.js).

The open-source innovation in interactive and data-driven storytelling, driven by news outlets like ProPublica and NPR as well as by a community of independent developers and agencies, fuels advancements in interactive data narrative and fosters greater public engagement with iterative development processes.

### 3.5.4 Responsive story output

Additionally, the field survey reinforced the importance of strong performance on mobile devices. Some of the projects analyzed performed well on mobile devices, with content loading quickly and adapting to screen size (responsive design), while others barely loaded on mobile devices. All of the projects were analyzed with Google Page Speed Insights, which highlights load time and mobile user experience (1 – 100). User experience ratings are based on multiple factors, including plugins, legible font sizes, and button size. Mobile user experience ratings ranged from very low (50s) to high (90s). Guided tours generally ranked high on mobile user experience, with many scrolling stories ranking low. The low rating of scrolling stories might derive from multiple
content types on a single page, necessitating all content to load upon initial site visit. Given the prevalence of mobile devices, online interactive stories should be responsive and optimized for mobile devices; the design intervention proposed in chapter four accounts for presenting story outputs on a range of screen sizes.

3.5.5 Lower technical and financial barrier to entry

Of the projects analyzed, many necessitated significant financial resources and technical expertise. ProPublica’s *Louisiana’s Moonshot*, for example, illustrates an innovative approach to sourcing data and translating that data into online narrative, but necessitated significant resources. Sourcing data via balloon and kite mapping required time, financial resources, and the willingness to travel. Furthermore, weaving together various data sources to create interactive maps required significant technical expertise (matching various map layers and orthorectified imagery requires a suite of tools and specialized knowledge). Similarly, *Planet Money Makes a T-Shirt* was only possible because of the significant public support via the Kickstarter campaign. The strong user experience of *Planet Money Makes a T-Shirt* derives from the thorough design process and extensive user testing. *Planet Money Makes a T-Shirt* is exceptional even within the context of NPR Visuals.

The immense resources necessary to create interactive stories such as *Louisiana’s Moonshot* and *Planet Money Makes a T-Shirt* preclude most individuals and organizations from producing similarly structured content. While such stories could never be replicated purely with out-of-the-box tools, user-friendly tools and resources could lower the barrier of entry for creating interactive stories from data.
3.6 Foundation for CartoStory

This chapter analyzed interactive storytelling projects incorporating geographic or geolocated data, including stories from established news outlets adapting to online innovation and independent producers exploring creative online presentation formats. The survey highlighted several structural and design strategies, as well as newly emerging approaches to sourcing data and presenting data online. The projects demonstrate a rising interest in creating interactive online stories that interweave quantitative (in this case, geolocated) data with qualitative content; reveal design strategies spearheaded by news organizations, interactive agencies, and independent producers; and demonstrate the importance of lowered barrier to entry to creating interactive data stories. These insights are foundational for design and functionality decisions for CartoStory, the tool discussed in the next chapter (chapter four).
Chapter four

CartoStory: Design intervention

Building upon the insights from the field survey, this chapter outlines a tool for adding a narrative layer to geolocated, citizen-collected data. Thus far, this thesis reveals the need for tools that provide greater integration of geolocated data into other narrative content and have a low technical barrier to entry. This section outlines the initial design process for CartoStory, a storytelling tool that addresses these dual needs. CartoStory allows users to create short, narrative tours of map-based data. Using a map base layer, users can choose particular data points, provide guided context through those data points, and output a shareable story. The goal of CartoStory is to enable the creation of compelling, narrative experiences from map-based data that are accessible to wide public audiences. This chapter focuses primarily on the design process behind CartoStory; while initial prototyping is underway, this design and research process is primarily intended to outline the optimal design and functionality of the tool. Beginning with a discussion of the impetus behind CartoStory, this chapter outlines the initial design process, including evaluating existing tools and creating wireframes and designs. This chapter also outlines target users and potential user flows and ends with a discussion of future modifications.
4.1 Impetus

The field survey illustrated a range of innovative approaches to crafting narratives from data, including novel approaches to acquiring data and presenting data online. The projects surveyed, however, illustrate opportunity for greater integration of geographic data and narrative to compel audiences to more deeply engage with data. Additionally, chapters two and three suggest that tools with low technical barrier to entry would allow individuals and organizations to leverage increasingly available data. As illuminated in the field survey, many innovative stories necessitate significant financial resources and technical expertise. ProPublica’s *Louisiana’s Moonshot* story required project producers to travel to Louisiana to source aerial imagery and then use tools to orthorectify imagery and translate remote sensing data sources into navigable maps. Likewise, *Planet Money Makes a T-Shirt* was only possible because of significant funding from NPR Planet Money’s highly successful Kickstarter campaign. Many projects in the field survey use a combination of open-source Javascript libraries, multimedia software, and custom design to craft online narratives; creating interactive data stories remains an often laborious and technically rigorous process.

Tools for data visualization and interactive storytelling, however, are gradually becoming more accessible to users without coding and design experience. For example, users with minimal technical skill can translate geolocated data into annotated maps using Google Map Engine or Mapbox’s Tilemill, or create a data visualization using Tableau or Google Fusion Tables. In Nieman Lab’s 2015 “Predictions for Journalism,” Alberto Cairo stated that, “In more and more publications, we see people who are not necessarily visualization designers or professional infographics designers who are creating visualizations” (2014).
Creating highly custom stories, though, remains a resource-intensive process. News agencies and independent producers alike recognize the importance of streamlining and reducing overhead during design and development processes. The 2014 leaked New York Times innovation report includes a section titled, “Balancing Act: One-offs vs. Replicability” where they describe the importance of replicability and “tools and templates.”

We have a tendency to pour resources into big one-time projects and work through the one-time fixes needed to create them and overlook the less glamorous work of creating tools, templates, and permanent fixes that cumulatively can have a bigger impact by saving our digital journalists time and evaluating the whole report. We greatly undervalue replicability...That runs counter to the approach at so many of our digital competitors. “We are focused on building tools to create Snowfalls every day, and getting them as close to reporters as possible,” said Kevin Delaney, editor of Quartz, which is known for innovative storytelling formats. “I’d rather have a Snowfall builder than a Snowfall” (2014, 36).

While The New York Times is a major news organization, independent producers and the general public face similar challenges. “Tools, templates, and permanent fixes” make it easier to craft stories without immense financial and technical resources. CartoStory addresses this need through enabling users to easily add a narrative layer to geolocated, citizen-collected data, thereby translating data into stories rather than simply maps or data visualizations. CartoStory is useful for linking data collection initiatives with visual storytelling capacity, but is also relevant for journalists and independent producers.

4.2 Design process

Initial design of CartoStory involved a user-centered design process drawing from the insights and opportunities highlighted in chapters one through three. The design process began
with analyzing the affordances and limitations of existing tools, including creating stories using some of these tools. Analyzing existing tools yielded insights for minimizing technical complexity; improving the narrative structure of outputted stories; and integrating data with text, photos, and other content.

The next phase involved identifying target users and outlining the core needs of those users. After establishing target users, the design process moved toward outlining core functionality; creating wireframes and user flow charts; and designing mockups of the homepage and other key pages. The next phase involved initial prototyping, including translating the designs into front-end code and initial application development using meteor.js. The flowchart in Figure 1.1 below illustrates the first steps of the design process, including comparative analysis, wireframing, and creating mock stories. While development of CartoStory is still underway, the design process outlined in this thesis lays the groundwork for core design and functionality.
Fig. 4.1: Flowchart illustrating initial phases of the design process

- Odyssey.js analysis
- Storymap.js analysis
- Meograph analysis
- Google Earth Tour Builder analysis

Wireframes for map-based storytelling tool

Mock stories with wireframes
4.2.1 Comparative analysis

The comparative analysis included evaluating Odyssey.js, Storymap.js, Meograph, and Google Earth Tour Builder. Evaluating these tools revealed opportunities for improved design patterns, data integration strategies, and lowered technical barrier to entry. Additionally, the development of NetStories, a database of online storytelling tools, broadly informed the analysis of existing tools.\(^{25}\)

Fig. 4.2: The NetStories online tool database

\(^{25}\) I created the NetStories tool site in collaboration with a team from the Center for Civic Media and Harvard’s Berkman Center for Internet and Society.
The NetStories project involved evaluating a range of online storytelling tools, including tools for mapping, data scraping, data visualization, and interactive video. Tool evaluation included difficulty level, narrative strategy, cost, inputs (e.g. photo, video, text), and outputs (e.g. map, interactive video). Insights from the NetStories tool evaluation and taxonomy broadly informed functionality choices during the CartoStory design process.

Fig 4.3: An example of individual tool display on NetStories

The comparative analysis included Odyssey.js, a map-based storytelling tool created by CartoDB and Vizzuality. Odyssey.js allows users to create narratives from a CartoDB base layer. While Odyssey.js is a Javascript library, the tool includes a “Sandbox” for creating stories without extensive coding. Using the sandbox, users can edit markdown to add specific data points based on geo-coordinates and add text and photos to those data points. Users can choose from three
different output formats, including a slide format where each slide contains content associated with a specific, geolocated data point.

While Odyssey.js is a unique approach to crafting narratives from a map base layer and reflects CartoDB’s innovation in online mapping, there are limitations to the story building approach employed in Odyssey.js. Although the Sandbox eases the process of creating stories, the Sandbox still requires users to edit markdown, which could pose a challenge to users unfamiliar with markdown. Furthermore, the story outputs display only a single data point at a time; for example, in the slide format, clicking through slides reveals a single geolocated data point and associated content. For map-based data with significant proximity significance, this structure eliminates the contextual perspective of a data point in relation to other points. Geographic patterns, such as points along a bus route or clustered in a specific neighborhood, are difficult to ascertain without viewing multiple geolocated data points simultaneously. Despite these drawbacks, Odyssey.js represents significant innovation in map-based storytelling.

Storymap.js is a tool that allows users to create map-based stories outputted as interactive slideshows. The outputted stories overlay large slides, containing images and text, on a base map. A subtle visual line connects data points, allowing audiences to notice geographic patterns. Creating stories with Storymap.js involves adding content through Storymap’s GUI26, which connects to a user’s Google Drive account. Users can add slides with headlines, descriptive text, and media (photos and videos) uploaded from the web or from the user’s computer. The interface includes two modes, edit and preview, allowing users to easily toggle between modes and preview stories.

---

26 Graphical User Interface
Storymap.js's strength lies in the visual linkage between map markers and the large, digestible slides. One drawback to the Storymap.js structure is that audiences must click through slides to view content. As discussed in chapter three, minimal user agency often provides a highly engaging user experience. Offering users a path to explore a story without clicking multiple times, such as a scrolling story or a guided tour, may engage users who would otherwise not click through a story. Furthermore, highly varying zoom levels and significant geographic shifts in outputted stories could be disorienting to audiences.

Meograph is a tool launched in 2012. Originally designed to create interactive videos featuring maps, photos, and other contextual content, Meograph has since shifted orientation and dramatically modified its user interface and story output type. Reviewing Meograph’s original premise and early example story outputs informed the comparative analysis. According to an article posted on PBS Idea Lab, the original goal of Meograph was to let you “create and share interactive stories that combine video with maps, a timeline and links, filling in that often missing context of where and when” (Lasica 2012). The article describes the functionality as follows:

When you create an account, you log in to create a story in as little as 10 minutes by following simple prompts in an intuitive interface. Then you add media — any YouTube video, along with photos, maps, time stamps, narration and annotations — to give it that 4D storytelling flavor: content, context, interactivity and connection (and links to explore at a deeper level) (Lasica 2012).

Audiences watch Meograph stories unfold without necessarily interacting, almost like watching a video (similar to video, Meograph stories progress through a timeline). For example, an outputted story could include a map background overlaid with rotating images and text, a format that interweaves time-based and location-based events into a coherent, highly guided
narrative. While the Meograph structure offers a highly guided experience, videos, photos, and text content are not necessarily integrated into the map (stories may play primarily as a series of photos and YouTube videos). Meograph nonetheless illustrates potential for integrating various data sources into a guided narrative experience resembling a video.

Google Earth Tour Builder offers users a means of creating 3D interactive tours using Google Earth. According to Google, “Tour Builder is a new way to show people the places you’ve visited and the experiences you had along the way using Google Earth. It lets you pick the locations right on the map, add in photos, text, and video, and then share your creation” (Google 2015). To create stories using Google Earth Tour Builder, users install the Google Earth Chrome browser plug-in. The builder interface is intuitive; users add slides to specific geolocated points. One drawback to Google Earth Tour Builder is the significant processing power necessary to run Google Earth map layers in a browser and the necessity of installing browser plugins to use the tool. The story output format also requires users to click through each slide. Google Earth Tour Builder, however, reflects significant innovation in 3D interactive map-based story building.

The comparative analysis yielded insight into the affordances and limitations of existing tools and potential for improved user interface and output format. Insights include need for highly guided narrative structure that allows users to experience a story with minimal clicks; built-in narrative arc; and minimal required technical expertise. The comparative analysis, along with the field survey, provide the foundation for CartoStory design and functionality choices.
4.2.2 Design and functionality

The design and functionality of CartoStory allows users to craft online narratives through the following steps:

1. **Importing map base layer or manually choosing data points:** Users can import data already aggregated on a map base layer (i.e. for citizen data collection initiatives) or manually add data points to a map. Users can add a base layer by entering the URL for an existing cartoDB map base layer. After entering the CartoDB URL, the map base layer automatically populates the story builder interface. Users can also manually choose data points by double clicking on specific points on the map, which automatically generates a data point at that location.

2. **Adding a title and introduction:** Users are automatically prompted with three initial “slides,” including an introduction slide, a data point slide, and a conclusion slide. Users can edit the title and introductory content directly in the slide interface (the initial content is placeholder content directing users to edit the title and introductory text). Content is directly editable, allowing users to edit text and view changes in real time without having to enter a “preview” mode.

1. **Adding context to data points (photos, text, and videos):** Users can add content associated with particular data points by clicking on an “Add Slide” icon and then selecting a specific geolocated data point on which to associate content. After adding a slide, users can add a headline, photos, and descriptive text to the slide. If users imported
a CartoDB map base layer with geolocated data, data associated with a particular geolocated point automatically populates the slide. In cases where significant data is associated with each geolocated data point, the first associated image and first two associated data fields are imported into the slide. Users can select from data already associated with a geolocated data point, or add additional photos and text from the web or uploaded from their computer.

2. **Adding conclusion and / or call to action:** By default, every story building interface includes a conclusion / call to action slide. This slide contains placeholder text that users can edit to add a conclusion. Depending on the goal of the story, users can add a call to action button to link to another URL (e.g. a web site or online petition). Including an explicit slide for conclusion / call to action encourages users to create a coherent narrative arc and engages users with a means of becoming more involved. If the story links with a citizen data collection campaign, the call to action could prompt users to contribute data to the story, thereby providing a link between citizen data collectors and audiences.

3. **Exporting and sharing story:** After finishing the story building process, users can export and share their story online. Exporting generates embed code and a live URL for viewing online. Every story includes social sharing links (Facebook and Twitter) to encourage audiences to share stories via their social networks.
All steps in the story building process lie primarily in a single user interface. Figure 4.4 is a design mockup of the user interface, including geolocated data points and slides containing content associated with specific data points.

Fig. 4.4: Initial design mockup
The left column features slide icons of core story components (the image above includes an introduction slide, two data point slides, and a call to action slide). These slide icons allow users to add slides, re-order slides, and select and edit specific slides. Upon clicking the "add" icon, users are prompted to choose an existing geolocated data point or add a geolocated data point to the map.

The outputted story format is a scrolling story. Drawing from insights from the field survey and analysis of user agency, the selection of scrolling output format allows users to engage with minimal required interaction. Scrolling guides users through content, triggering the appearance of specific slide content and highlighting the map markers associated with that slide. All geolocated data points are simultaneously visible, so users can contextualize specific data points and recognize geographic patterns. A subtle graphical line links data points in the specific order designated by the narrative, thereby providing a visual cue for the narrative arc.
4.2.3 Differentiators

The design and functionality of CartoStory are based on the following priorities and differentiators:

*Built-in narrative arc:* CartoStory encourages users to provide a guided, narrative path through data points. Building in a narrative arc is accomplished through providing introduction and conclusion slides and including placeholder content that guides users in choosing content. Furthermore, the interface could include tooltips to provide additional guidance for narrative
structure. Many storytelling tools—including those featured in the comparative analysis—do not guide users through the process of crafting narrative. Providing guidance for constructing narrative results in more engaging, user-friendly story outputs.

*Highly guided narrative experience:* The output format is a highly guided scrolling story. As illuminated in the discussion of structural strategies and design patterns in chapter three, scrolling stories provide a highly guided narrative experience and are increasingly used by news organizations and independent agencies to present interactive and data-driven content. Among map-based scrolling stories, the strongest examples evaluated in the field survey integrate dynamic data visualizations triggered by scrolling action, such as *Faces of Fracking* and *The New York Times’ The New Silk Road*. Among the tools evaluated in the comparative analysis, most require audiences to manually navigate through stories (e.g. clicking through slides). CartoStory does not require users to manually navigate outputted stories (besides scrolling) and maintains content as a cohesive narrative arc rather than breaking up content into successive slides.

*Low technical barrier to entry:* Creating online narratives using CartoStory does not require technical knowledge or coding experience. Tools like Odyssey.js have a relatively low technical barrier to entry, but Odyssey.js still requires users to edit markdown to create stories using the Sandbox, which limits the range of potential users. To truly democratize the process of creating interactive map-based stories, users should be able to create stories with no technical knowledge. CartoStory allows users to create stories by manually adding data points and content without coding or editing markdown. Furthermore, the tool is open-source and accessible to anyone with internet and computer access. Most of the projects analyzed in the field survey and case studies
involved significant custom design and development; CartoStory allows users to create compelling interactive narratives, with a small degree of customizability, without prior design and development experience.

*Linkage with citizen data collection:* CartoStory allows users to import a map base layer with aggregated data points. Thus, if users are involved in a data collection initiative and have aggregated data using a platform like CartoDB, they can easily import this as a base map layer and add context to specific data points. Using a data collection tool such as Open Data Kit, a community of citizen data collectors could collect data in their community on issues such as water quality or public infrastructure, and add a narrative layer to that data.

*Responsive:* In the field survey and comparative tool analysis, only some outputted stories respond to screen size. The output format of CartoStory will be responsive, allowing users to easily view stories on mobile phones and varied devices. While the story building interface is designed for use on a desktop or laptop, the output story format is fully responsive and allows users to share with wide networks of people using various devices.

### 4.2.4 Technical implementation and initial prototyping

After designing and specifying the functionality of CartoStory, an initial prototype is under development. Rather than focusing on building and deploying CartoStory, this thesis is primarily focused on analyzing interactive data storytelling strategies and outlining the proposed design and functionality of CartoStory. Nonetheless, some initial prototyping is underway and provides a stronger foundation for design and functionality decisions.
The prototype is being developed using Meteor, a Javascript web application framework. Meteor streamlines the application development process through packaging dependencies and enabling reactive development so code changes are presented client-side in real time (Meteor 2015). The front end of the site is built with Bootstrap, a framework for front-end web development using HTML, CSS, and Javascript (the initial Photoshop designs were translated into front-end code using Bootstrap). By default, Bootstrap is responsive, so the CartoStory builder interface can adapt to various screen sizes (although the tool is not designed for creating stories on a mobile phone). The image below is an early iteration of the front-end coded using Bootstrap. Meteor has Bootstrap core files built-in, although a new CSS file was added to customize the styling of the site beyond Bootstrap’s default styles.

Fig. 4.6: Early front-end of CartoStory created using Bootstrap with map base layer (Stamen) using Leaflet.js as a web mapping service
For prototyping, the map base layer is Leaflet.js, an open-source Javascript library for creating online interactive maps. Leaflet serves as the Web Mapping Service (WMS); map data from services including CartoDB, Mapbox, and OpenStreetMap can be loaded atop Leaflet (Leaflet supports GeoJSON layers). To integrate Leaflet.js into Meteor, an open-source package called “Leaflet for Meteor” was integrated into Meteor. With this package, double clicking on the map base layer adds a map marker. During development, the tile layer provider was specified as Stamen Watercolor, a map base layer created by Stamen Design.

Account capabilities are being integrated into CartoStory, enabling users to log in, add custom content, save their work, and share stories. To implement user accounts, an open-source package called “User Accounts for Meteor” was added to Meteor, allowing for user interface templates based on Bootstrap and fully functional user account creation capabilities. “User Accounts for Meteor” provides stylistic customizability but also provides core templates. The package enables users to sign in through a custom account or through an existing Facebook or Google account (CartoStory only features custom account creation).
Design customization included adding a custom Google Font powered by the Google Fonts API; using a custom color palette based on the blue color used in preliminary designs; and integrating icons downloaded from open icon libraries.

4.2.5 User types and scenarios

CartoStory caters to two primary use cases and user types. The first potential user base are citizen data collectors focused on a specific campaign, organization, or issue. The second user base are journalists or producers who want to create online narratives based on geographic content.
**Citizen data collectors:** Citizen data collectors could range from a single individual or a distributed network of people collecting data on a particular issue. On a local level, a neighborhood group could collect data on building accessibility by using a mobile data collection app to capture photos and geocoordinates of ramps and entrances. On a larger scale, a distributed network of individuals around the world could collect data on water quality using the Open Water Project’s Riffle. Citizen data collectors could use data for advocacy (e.g. organizing for building accessibility) or for journalistic and educational purposes (e.g. telling a story about a local environmental dispute). Aggregated data can be imported into a map base layer using a platform such as CartoDB (CartoDB accepts a range of file formats, including CSV, KML, XLS, and GeoJSON), and then loaded into CartoStory to create an interactive narrative.

**Journalists and producers:** Journalists and independent producers would likely create a story without aggregated, citizen-collected data. They could import an existing dataset into CartoDB and use this as a map base layer, or manually add data points to the default map base layer. Independent journalists, news agencies, design agencies, and others interested in crafting map-based stories could use CartoStory through manually adding data points. For example, a journalist could create a map-based story illustrating urban sprawl or shifting coastline conditions in a particular region.
Fig. 4.8: User flow for two potential user types: citizen data collectors (user flow 1) or journalists and independent producers (user flow 2)

**User flow 1:**
**Citizen Data Collectors**

1. Citizens collect data (e.g. with mobile data collection tool)
2. Aggregate data on map base layer
3. Import map base layer into CartoStory
4. Choose from content associated with geolocated data points
5. Add additional narrative content
6. Add conclusion and optional call to action (e.g. share story)
7. Export story and share via social media and direct link
8. Viewers engage with issue and potentially collect data

**User flow 2:**
**Journalists & Producers**

1. Add data point to map
2. Add narrative text and other content (upload or select from web)
3. Add conclusion and optional call to action (e.g. share story)
4. Export story and share via social media and direct link
5. Viewers learn about issues, potentially share story
**User Flow 1 (citizen data collectors):** The diagram on the left represents user flow for citizen data collectors. This user type starts with an issue (e.g. air quality or urban infrastructure) and then crowdsources citizen-collected data on that issue. After aggregating data on a map base layer, they import the base layer into CartoStory. Next, users can chose from content associated with geolocated data points (data associated with specific geolocated points is available for story creators to draw from) and can add additional narrative content including text and images. Users can add a conclusion and optionally a call to action; the call to action could range from “share story” to “contribute to data collection” to “get involved by attending a city council meeting.” Users export the story and share through social media channels (Facebook and Twitter) or through a direct link. Audiences learn about the issue and potentially get involved through contributing to data collection.

**User Flow 2 (journalists and producers):** The diagram on the right represents user flow for journalists and producers. This user type starts with a story suited to map-based narrative, and manually adds data points to the map base layer (they could also import a CartoDB map base layer with existing data points). Next, they add narrative text and other content by uploading content from their computer or selecting photos and content from the web. The user exports the story and shares through social media channels (Facebook and Twitter) or through a direct link. Audiences then learn about the story and potentially share the story through social media or direct link.
4.3 Future modifications

The prototype is at an early phase, and more development is necessary to carry out a field test. Further development could include designing and building a public “Stories” page that features a selection of recently created stories, allowing audiences to view previews of featured stories and click on a story to view the full content. A robust “Stories” page would prompt audiences to learn about particular issues and potentially motivate them to create their own stories.

Additionally, a future iteration could include both the scrolling story and guided tour output formats. As suggested in chapter three’s field survey, both of these formats provided highly guided narrative experiences. Integrating a guided tour output format would involve significant design and functionality additions, including voiceover-recording capability, but this format would strengthen the customizability and narrative capacity of the tool.
Chapter 5

Conclusions and future work

This research investigated interactive online narrative as a mode of communicating data and proposed a design intervention for adding a narrative layer to geolocated, citizen-collected data. A discussion of today’s proliferation of data and innovation in interactive storytelling laid the foundation for analyzing the field, and an analysis of twenty-five projects from news organizations, interactive agencies, and independent producers led to insights regarding design patterns, data sources, and structural strategies. Analyzing specific projects surfaced three structural strategies—the guided tour, scrolling story, and interactive web documentary—as well as novel approaches to collecting and sourcing data. The insights from this analysis laid the groundwork for a design intervention called CartoStory, a tool that allows users to create short, narrative tours from map-based data.

This chapter addresses the implications of innovation in interactive data storytelling, focusing on subjectivity and advocacy efforts as well as public engagement with complex data. This chapter also suggests potential opportunities for further investigation and design innovation, including streaming real time data, physical user engagement, and balancing ease-of-use and customizability in tool development.
5.1 Implications of innovation in interactive data narrative

While this thesis focuses primarily on design and functionality of tools and projects that weave data into interactive online narrative, there are vast implications of this mode of communication. Translating data into interactive online narratives allows for advocacy-oriented and journalistic stories that are easier to digest than “raw” data; narrative allows for a subjective, authorial stance that ultimately behooves creators who want to highlight particular insights from data. Furthermore, interactive narratives foster public engagement with complex data on issues that are difficult to visually represent, such as shifting air quality or coastal erosion.

5.1.1 Advocacy and intentionality

Crafting narratives from geolocated data is a subjective process with significant authorial agency. The mapmaking process itself is imbued with socio-political significance, as suggested by critical cartographers (Crampton and Krygier 2006); choosing data sources, map projections, and visual styling are embedded in the politics of representation. In adding a narrative layer to geolocated data, technical and aesthetic choices such as code libraries, user experience, and narrative arc extend the subjectivity inherent in mapmaking. This research focuses primarily on strategies for crafting interactive narratives from data, but underlying these strategies is the subjectivity and authorial voice implicit in narrative.

The projects evaluated in the field survey approach subject matter from various angles, including advocacy, journalism, and citizen engagement; many of the projects reflect the unique methodological approach and epistemological stance of their umbrella organization. The goal of NPR Visuals, for example, is to evoke empathy in their audience, and the Planet Money Makes a
T-Shirt web documentary accomplishes this goal through weaving personal stories into data visualizations and other content. On Brian Boyer’s personal website, he describes the importance of empathy:

We’re here to create empathy. To introduce you to somebody you’ve never met, and think for a few minutes about life in their shoes...And we’re here to serve our audience, so yes, pageviews or it didn’t happen. And I think visuals have the power to deliver audience to a difficult story. But our success is measured in engagement, not views. User-centered design, not interactive gimmicks. Affecting stories, not clickbait (2014).

ProPublica, the news organization behind the Losing Ground and Louisiana’s Moonshot stories, describes itself as producing “investigative journalism in the public interest. Our work focuses exclusively on truly important stories, stories with 'moral force”’ (ProPublica 2015). The subject matter (industrial activity and environmental restoration) and data sources (government and independently sourced remote sensing data) of Losing Ground and Louisiana’s Moonshot reflect ProPublica’s public interest goals.

Crafting interactive narratives reinforces the underlying goals of sourcing and presenting data, whether driven by a public interest news organization or an advocacy-oriented community group. Through the process of crafting data narratives, journalists, producers, and community organizations tell stories from a particular, and intentional, stance.

5.1.2 Public engagement with complex data

In addition to advocacy, crafting narratives from geolocated data enables deeper public engagement with complex datasets. Drawing insights from “raw” aggregated data is difficult, whether a spreadsheet, XML document, or other format; even with a high degree of data literacy, an individual may need hours of analysis to draw insights. Translating aggregated data into a
data visualization or annotated map allows users to engage more easily with data, but may still require in-depth navigation to draw insights. For instance, faced with dozens of annotated map markers, how does a user choose where to explore? In contrast with “raw” data, annotated maps, or data visualizations, interactive narratives shift data closer to digestible, accessible formats. Rather than facing endless options for exploration, narratives provide guided context through data.

Furthermore, interactive data narratives enable visual depiction of issues that are otherwise challenging to represent. For example, creating narratives from data on the environment and common resources allow audiences to visually digest environmental shifts occurring over significant time or geographic scales. Data narratives can strengthen public understanding of a range of issues (e.g. environmental, urban, governmental), but are especially relevant to environmental representation due to the challenges facing environmental communication. Early studies in environmental communication illustrate the divide between the rapid pace of the news-cycle and the slow pace of environmental change (Hansen and Machin 2013; Schoenfeld, Meier, and Griffin 1979). Environmental shifts happen over long durations (often thousands of years) and environmental challenges, such as degrading air quality, are frequently invisible to the eye. Given these constraints, media outlets face difficulty accurately representing the environment. Other studies reveal the dissemination of romanticized images of the environment and the tendency of media outlets to distill complex environmental processes into a single concept or image. According to Hansen and Machin, “Visual representations of the environment tend to be decontextualized and aestheticized in ways that enhance their flexible and versatile use across different genres of communication” (2013, 157). Translating environmental data into interactive narratives addresses some of the challenges of environmental
communication through illustrating shifts over large time and geographic scales and visually representing arcane data. The example of ProPublica’s *Losing Ground* interactive – focusing on industrial activity and coastal erosion in Louisiana over the last century – illustrates the potential of interactive data narratives to address significant environmental changes and promote public engagement with complex data.

### 5.2 Future opportunities

This research evaluated strategies for crafting narratives from data, highlighting techniques that interweave geolocated data with narrative text and other content; many of these strategies informed design and functionality decisions for CartoStory. Some emerging techniques, however, were not included in the scope of this thesis but portend future directions for interactive data narrative. Opportunities for innovation include streaming real time data, physical user engagement, and balancing ease-of-use and customizability in tool development.

#### 5.2.1 Streaming real time data

Most interactive narratives draw from previously aggregated data rather than streaming real time data. Indeed, CartoStory is designed for data already aggregated on a map base layer or manually added during the story building process. Streaming real time data in interactive narratives, however, is an area of immense opportunity.

Technical constraints to real time data visualization are rapidly diminishing, opening doors for innovation in real time interactive data journalism. In October 2013, CartoDB launched a feature called “synced tables,” allowing users to feed data from “any supported file
format available on the web and have CartoDB sync to that data keeping it real-time forever” (CartoDB 2013). Applications of synced tables include real time updates to data visualized from a manually updated Google spreadsheet or the real-time display of continually updated government weather data.\textsuperscript{27}

Furthermore, the New York Times R&D department recently launched StreamTools as an alternative means of interrogating data, enabling data analysts to work with a continual stream rather than awaiting aggregation in a database (Dewar 2014). A post on The New York Times R&D Lab’s blog contextualizes Stream within the R&D Lab’s forward-looking approach: “Like all projects at The Time’s R&D Lab, Streamtools is designed to probe a particular future approximately 3 to 5 years away” (Dewar 2014). Streamtools,

provides a general purpose, graphical tool for dealing with streams of data. It offers a vocabulary of operations that can be connected together to create live data processing systems without the need for programming or complicated infrastructure. These systems are assembled using a visual interface that affords both immediate understanding and live manipulation of the system (Dewar 2014).

Tools like CartoDB’s Synced Tables and New York Times R&D Lab’s Streamtools address the prescient question of how to incorporate live sensor data into interrogation, analysis, and visualization of data.\textsuperscript{28}

One example of an interactive story drawing from real time data is \textit{Into the Okavango}, a piece produced under National Geographic that allows users to follow a “live data expedition”

\textsuperscript{27} Real time data visualizations are generally synced at particular intervals (e.g. every one hour or every three hours) rather than livestreaming data in real time.

\textsuperscript{28} While CartoBD’s synced tables reflects a shift toward democratizing real time data visualization (i.e. low technical barrier), The New York Times R&D Lab explores related questions albeit aimed at a developer communities. The New York Times R&D Lab’s blog suggests that StreamTools addresses a future shift toward data provided as streams and the unnecessary presence of databases “between the sensor and the data scientist” (Dewar 2014).
involving a team of collaborators traveling down the Okavango Delta, a wetlands wilderness in Botswana: “Into the Okavango displays data which was uploaded daily, via satellite, by the expedition team in the Okavango Delta. Data is also available through a public API, allowing anyone to re-mix, analyze, or visualize the collected information. The software allows viewers to virtually experience one of the last remaining areas of true wilderness, normally inaccessible to the average person” (The Office for Creative Research 2013). The underlying goal of the “live data expedition” is to raise awareness about the environment and “fragility” of the Okavango wilderness area. Into the Okavango explicitly links open and accessible data with increased capacity to understand the natural environment: “Remote sensing technologies, open APIs, Apps that enable digital data collection and real-time sharing, increased access to mobile connectivity, and the rapid growth of online citizen science web portals like eBird are the first signs of the tidal wave of development and investment in a global open access biodiversity and ecosystems monitoring network geared at prioritising conservation action” (Boyes 2014).

Streaming “live data” – as illustrated in the Into the Okavango project – engages audiences in an unfolding journey and provides incentive to return to an interactive experience. Despite the increasing availability of tools enabling real time data streams, however, most live data visualizations incorporate minimal narrative elements. Generally, live data visualizations function similarly to static data visualization but with a regularly updated data feed. Given the high rate of progress in real time data streaming and decreasing technical barriers to participation, however, more interactive narratives will likely integrate real time data in coming years.
5.2.2 Physical user engagement

Another area of opportunity is interactive stories involving physical user engagement. While interactive online stories are usually constrained to screens (whether desktop, laptop, or mobile device) with users engaging through a mouse or touch screen, users could participate through spatial and physical engagement. For example, Up in the Air – a project created at the Datalore Hackathon at the MIT Media Lab – engages users beyond simply mouse clicks and screen swipes. Up in the Air began as a project to explore atmospheric data, and manifested as an interactive online map where users breathe air into their computer to interact with the story.29 The users' breath visually manifests on-screen and travels around the world triggering imagery of climate related events associated with specific countries. Up in the Air actively engage users in a novel form of participation (the physicality of breathing) and suggests potential for physical engagement with interactive data stories.

5.2.3 Balancing customizability and technical accessibility

One of the issues addressed in this research is the technical barrier to entry to creating interactive stories based on data. Often, technical skill required to use a tool correlates with customizability. Many newly available tools lower the technical barrier to entry, but those tools often output templated stories with minimal customizability. Crafting stories through custom design and coding maximizes customizability (using Javascript libraries such as Popcorn.js or Odyssey.js eases the development process for implementing particular features). With drag-and-

29 The computer's microphone senses the user's breathing.
drop tools such as Scrollkit, however, users can create interactive stories without any coding experience; the drawback to such tools is limited ability to customize the aesthetics and functionality of outputted stories. Future advancements in tool design and functionality will likely diminish the one-to-one correspondence between difficulty level and customizability and enable highly customized narrative outputs.

5.3 Conclusion

This chapter discussed implications of crafting interactive narratives from data, including supporting advocacy campaigns and encouraging public engagement with complex data. This chapter also suggested potential areas of additional research and design innovation, including streaming real time data, physical user engagement, and tools that balance ease-of-use and customizability. The next several years will likely witness immense innovation in strategies for collecting data and creating interactive narratives from data, including strategies that transcend the screen-based and browser-based stories discussed in this thesis.

As a whole, this thesis investigated the emerging field of interactive data storytelling, including analyzing the field for design trends and emerging strategies; laying the groundwork for CartoStory, a tool for adding a narrative layer to geolocated, citizen-collected data; and discussing the implications of this mode of communicating data. While this research outlines patterns and innovations in interactive data storytelling, it also suggests areas of exploration and innovation for the coming years.

30 Scrollkit was discussed in chapter three, "Design patterns and structural strategies."
Underlying almost every project examined in this thesis is a particular, authorial voice. Whether produced by a news organization, community group, independent producer, or design agency, creating interactive data narratives a subjective process. This question of voice and subjectivity in storytelling evoked debate at a workshop held at the Center for Civic Media in January of 2015 called “From Sensors to Stories.” The workshop brought together individuals from organizations working in data collection and visualization—including Ushahidi, Public Lab, Witness, and ProPublica—to discuss strategies for crafting narratives from data. During the discussion, one of the participants spoke up and ardently argued against adding narrative to map-based data on the basis that narrative adds a greater degree of bias and could undermine public confidence in that data. Of course, data is never truly objective: As described by Lisa Gitelman, “Like events imagined and enunciated against the continuity of time, data are imagined and enunciated against the seamlessness of phenomena. We call them up out of an otherwise undifferentiated blur” (2013, 3). The question of objectivity of data, however, was vehemently discussed at the “Sensors to Stories” workshop in the context of organizations telling stories with data.

Data are captured and represented through a process filled with human error and subjectivity. Yet, presenting data points on a map conjures the appearance of objectivity; map markers at specific geo-coordinates with succinct annotations appear to be accurate representations of reality. By using a tool like CartoStory to select data points, add and change content, and provide a guided path through those data points, human bias is more readily apparent. Tactical Technology Collective’s *Visualizing Information for Advocacy* addresses the

---

31 The workshop included a discussion of issues and challenges in collecting and visualizing data, a presentation of tools and trends in data storytelling, and a design exercise to quickly outline potential data storytelling tools.
trepidation associated with telling stories from data: “Combining narrative with facts to trace a storyline through data is difficult. Many advocates get nervous at the very idea of telling stories with data. It goes against the principles of working empirically, and we have to be careful not to fall into the trap of shoehorning facts into a pre-defined narrative” (2013). The benefit of adding a narrative layer to data, however, is that individuals and organizations can convey their perspectives and suggest particular insights. Perspectives and insights are central to interpreting data—whether journalistic, advocacy-oriented, or educational—and the design strategies in this thesis intentionally highlight those perspectives through online storytelling.
Appendix: Projects evaluated in field survey

Countries most exposed to the carbon bubble
http://www.theguardian.com/environment/interactive/2013/apr/19/countries-exposed-carbon-bubble-map

State of the Polar Bear
http://pbsg.npolar.no/en/dynamic/app/

State of the Salmon
http://www.stateofthesalmon.org/msc/

Bear71
http://bear71.nfb.ca/#/bear71

Disappearing Rio Grande
http://riogrande.texastribune.org

Faces of Fracking
http://www.facesoffracking.org/data-visualization/

The Downside of the Boom

The past, present, and future of CO2
iSeeChange
http://thealmanac.org/year.php

Up in the Air
http://upintheair.co

Forest Flip
http://forestflip.org

Into the Okavango
http://intotheokavango.org

What is Missing?
http://whatismissing.net/#/home

Planet Money Makes a T-Shirt
http://apps.npr.org/tshirt/#/title

Losing Ground
http://projects.propublica.org/louisiana/

Louisiana’s Moonshot
http://projects.propublica.org/larestoration

WNYC Cicada Tracker
http://project.wnyc.org/cicadas/

InfoAmazonia
http://infoamazonia.org/#/map=6731
Climate Commons
http://climatecommons.earthjournalism.net/map/

Ekuatorial
http://ekuatorial.com/en

Fraser Basin
http://ekuatorial.com/en

Return to Elwha
http://www.returntoelwha.com

Into the Okavango
http://intotheokavango.org/

Climate Change Threatens to Disrupt the Ranges of Birds

America’s First Climate Refugees
http://www.theguardian.com/environment/interactive/2013/may/13/newtok-alaska-climate-change-refugees
Bibliography


