# LOCAL ORIGINS

What does it mean to say, "All data are local"? Before proceeding to the main cases of the book, let me further unpack the title claim. It relies on two concepts, *data* and *local*, that are pervasive in contemporary popular and academic writing, but less commonly seen together. In order to explain their pairing for the purposes of my argument, I need to address a number of misconceptions about each term independently. In the process, I will take the opportunity to acknowledge the contributions of other scholars who have laid the substantial foundations for this work. Let me turn to conceptions of data first.

## TAKING DATA APART

### Data Is Not Singular

*Five-thirty-eight*, the go-to website in recent years for journalistic commentary on questions about data, proclaims, "It should be 'data is.' It just sounds wrong otherwise."[1] Anyone who has worked sufficiently with data has probably thought this or heard it announced by others. Although contemporary dictionaries and style guides maintain that both singular and plural uses are accepted (yes, "all data are local" is grammatically correct), popular writers and increasingly academics shy away from the plural.[2] Instead, they employ the term as a mass singular noun. After all, data seem ubiquitous, like water, air, or oil—to which data have sometimes been crudely compared as the next great resource. But those words have no singular. Data does: datum. Moreover, every decision about word choice cannot simply be about what sounds right, for that changes over time. If "data are" sounds funny, I ask the reader to embrace its strangeness, treating this as an opportunity to reflect on their use of the term more generally.

To emphasize the relative subtlety of this suggestion, consider that some scholars ask us to put aside the term *data* altogether. In order to foreground the human production of data, digital humanities theorist Johanna Drucker has a radical proposition: as opposed to saying *data*, which in Latin means "given," why not substitute *capta*, meaning "taken."[3] This would serve as a useful reminder; capta must come from somewhere and someone. I don't dispute Drucker's incisive point. Yet *data* is too widely used to be abandoned just now. We can, however, retain the plural usage of the term in order to reinforce a multifaceted perspective. Unlike the subjects of many mass singular nouns, we can take data apart (without the need for a chemical reaction). When we do so, we find that data are heterogeneous in ways that matter.

Throughout the book, I use *data* in its older plural form (unless I am referring to the word itself, which is of course singular). This acknowledges that data arise from and are used in varied circumstances worth acknowledging. Even the largest data sets

are often agglomerations of sources made in varying conditions—whether they are collected with different instruments, in particular organizations, or simply at distinct moments. While investigating the DPLA (profiled in chapter 3), which brings together an eclectic collection of digital artifacts and formats from across the United States, I was struck by the diversity of its formats. These differences can be attributed to many things. For instance, while archivists seek to determine and preserve the original or authentic ordering of a data collection, librarians attempt to bring each individual data point into an encompassing ordering system, such as the official Library of Congress notation, Dewey decimal classification, or more flexible Dublin Core, thereby facilitating access for a broader public.

Furthermore, we cannot speak about data independent of the plurality of encompassing structures through which they are collected, encoded, and managed: databases. These structures, typically hidden from view behind user interfaces, vary widely as well.[4] And like data themselves, databases are subject to material and historical constraints.

It is worth noting here what I mean by material constraints. I am building off the notion of "materialities of information representation" described by informatics scholar Paul Dourish. In his recent book *The Stuff of Bits*, Dourish writes at length about the "material forms in which digital data are represented and how these forms influence interpretation and lines of action."[5] So, for example, relational databases—the most common form today—are a 1970s' invention of IBM, meant to be data agnostic. They treat data as generic content, rigidly held in an array of rows and columns. Yet as critical cartography has taught us, there is nothing neutral about imposing a grid on the world.[6] Earlier databases were structured differently. Hierarchical, networked, and attribute-value systems are frequently overlooked alternatives in which data are more entangled with their database structure. Thus, the plural view of data should be extended to the ways in which data are stored, and in turn manipulated and retrieved.

## Data Are Not Universal

For the last few decades, scholars have developed empirical accounts of how data vary from one scientific or engineering context to the next.[7] These studies have largely sought to complicate a widely held but simplistic perspective: that data are universal and invariable. Science and technology studies (STS) scholar Bruno Latour deftly captured this purified conception of data with the term *inscription*. Latour explains inscriptions (i.e., data) as "objects" created for the production of scientific arguments, "which have the properties of being mobile but also immutable, presentable, readable and combinable with one another."[8] As a shorthand for this collection of features, Latour calls inscriptions "immutable mobiles."[9]

Many have challenged this overly abstract definition of data by exposing the ways in which data practices and data themselves differ from one context to the next. Research on the diversity of data has been conducted in the subdiscipline of laboratory studies,

which Latour helped to pioneer, as well as in museums, health care, space exploration, and climate debates, to name a few.[10] Outside of scientific and engineering work, studies of data reveal even further variation. Conceptions of data differ enormously within social and humanistic research—the latter of which has only recently adopted the term at all.[11] Each disciplinary community has its own rituals for determining when something should count as data. Moreover, in different arenas of public or civic life, data and infrastructures to support them are multiplying rapidly. In *The Data Revolution*, a comprehensive look at emergent practices with data, Rob Kitchin takes pains to account for all the sorts of data one might encounter today: "Data vary by form (qualitative or quantitative), structure (structured, semi-structured, or unstructured), source (captured, derived, exhaust, transient), producer (primary, secondary, tertiary), and type (indexical, attribute, metadata)."[12]

But accounting for the various genres of data does not help us understand their use and meaning. For as critical theorist Jonathan Culler reminds us, "Meaning is context-bound, and context is boundless."[13] Accounts of data too often leave out this broader context. Curiously, Latour's term *inscription*, translated from the Latin roughly as "write into," suggests that data are embedded, rather than autonomous.[14] Indeed, the term itself—if not Latour's usage of it—resonates with my view of data as enmeshed in sociotechnical contexts that shape their production and interpretation. How can we learn to engage with this encompassing context, which Kitchin, writing alongside Tracey Lauriault, calls a "data assemblage"?[15]

The Arnold Arboretum, profiled in chapter 2, illustrates how data are entangled within a knowledge system and inscribed in a place. For instance, the arboretum has long collected data on "provenance type," a classification used to distinguish plants by their origins: "W" marks a plant collected in the wild; "C" marks a plant sourced from a nursery or other institutionalized setting; "U" marks a plant of unknown origin; and "Z" singles out a plant as a cutting, taken from a wild plant that is already part of the arboretum collection. Knowing why provenance type matters to researchers at the arboretum, how it figures into the troubled history of the institution, and what about these designations incites controversy today is essential to understanding the arboretum's data along with the questions that they can and cannot help answer. Outside the context of their supporting knowledge system, such data are misleading at best.

## Data Are Never "Big"

*Big data* began as a buzzword, created by industry and embraced by popular media, in order to describe data sets of the 1990s and early 2000s so massive that they outgrew the existing tools to manage them.[16] More recently, big data has gained legitimacy as an independent area of research. Academics have since tried to formalize the term. Kitchin and coauthor Gavin McArdle define big data as those data that are principally high magnitude in volume, variety, and velocity. Yet a range of other characteristics are

often present in data sets described as big: exhaustive, high resolution, relational, and flexible.[17]

I have come to find, however, that more than anything else, *big data* is a term that speaks to our contemporary feelings about data. Social researcher Kate Crawford has called big data, data at the scale that inspires anxiety.[18] Her reference point for high anxiety data is the trove of intel collected from social media by the likes of the US National Security Agency or Britain's Government Communications Headquarters—purportedly to identify the next terrorist plot.

Those agencies have amassed enough data for all of us to feel surveilled, while also containing sufficient gaps to sow unshakable doubt among intelligence agents themselves. "We don't have context!" agonize the creators of the innocently named "Squeaky Dolphin" project, a recently exposed big data initiative created by the Government Communications Headquarters. Anxieties proliferate on all sides of big data. Crawford's definition is amusing and apt, particularly because "big" is relative. It can mean different things to different communities. Each group that makes use of data has its own capacities and thresholds for being overwhelmed by "the anxieties of big data."[19]

Despite these useful metacharacterizations of big data, all digital data are by definition agglomerations of small, discrete signals, represented as 0s and 1s in computer memory. Each byte, each item in a list, and each row or cell in a spreadsheet is part of a composite with a complex provenance. Whether generated by algorithms, created by instruments, or keyed in by catalogers, digital data have their own contingencies that are useful to understand. Frequently these differences, called "inhomogeneities" by historian of science and technology Paul Edwards, go unnoticed. Or if they are noted, they are filtered out in order to create the illusion of consistency, necessary to perform large scale queries or calculations with a data set.[20]

The term *big data* has endured beyond many expectations. Moreover, the ideology of big data has pervaded existing data initiatives that would not be considered big when judged by their scale alone. In examples such as the Arnold Arboretum, DPLA, News-Scape, and Zillow, the ideology of big data has infiltrated workaday practices with data sets that measure only in the tens of thousands or millions of entries. Aspiring to the ideology of big data means seeking to collect everything on a subject, downplaying the importance of data's origins, and assuming that data alone can entirely supplant other ways of knowing.[21]

Returning to the example of the DPLA, considered to be big data within the organization (signaling the library's anxiety about it), we can see that data from distributed collections across the country do not simply add up. As mentioned in the previous chapter, in one contributing collection alone, I found more than a thousand different date formats, from "ca. _ _ _ _ s" (of which there are 640 instances) to "probably before _ _ _ _" (of which there are only 7). Speaking about these data as "big" effaces the complexity of what initiatives such as the DPLA bring together.

## Data Are More than Rhetorical

Although I advocate returning to an older plural usage of *data*, my understanding of the term is not entirely old school. Indeed, I depart from historical understandings of data as primarily rhetorical. In its earliest applications as a theological expression in the seventeenth century, data meant either "the set of principles accepted as the basis of an argument" or "facts, particularly those taken from scripture."[22] In the late eighteenth century, the term was adapted to describe the results of empirical observations and experiments of the kind now associated with scientific practice. Even today, in varied realms of scholarship, data often means "alleged evidence."[23]

Information studies scholar Christine Borgman writes that "entities become data only when someone uses them as evidence of a phenomenon, and the same entities can be evidence of multiple phenomena."[24] She offers the example of excerpts from an old family photo album. One researcher might treat these as evidence of particular clothing styles common to an era. Another may use them to verify family relations. In other words, things become data within interpretative acts. Although this may hold true in some scholarship, I am focused on a broader spectrum of uses for data and their emergent social roles in the public realm. Today data are part of functional sociotechnical systems from which they cannot be easily separated. In this sense, data can be operational rather than rhetorical.

The operational function of data is most evident in the case of housing, explained in chapter 4. Zillow's interface draws together data from public and private sources across the United States in an effort to not simply represent the state of the housing market but also reshape its dynamics. The site offers visitors new estimates every day on homes across the country—their own or prospective purchases and rentals. In doing so, Zillow is effectively establishing a new subject position for buyers and sellers in real estate: seemingly empowered by numbers, but actually blinkered to the broader implications of their own market-frenzied choices for the future of affordable housing options. Data have become operational; they are enmeshed in the practices and politics of everyday life.

## Not Everything Can Become Data

In her rhetorical treatment of data, Borgman argues that they are defined not by "what" but rather "when."[25] Anything can become data, observes Borgman, if it is taken up as evidence in an argument, including texts, photographs, and even traces of pigment from an archaeological field site. Making data, though, is not simply an act of naming. In thinking about when something becomes data, we must lend weight to the material processes involved; measuring, recording, and otherwise capturing the world are processes with physical constraints. Making data means bringing a subject into a preexisting system, defined by durable conditions of data collection as well as storage, analysis, and dissemination.[26] Aspects of the original subject are inevitably lost in translation.

Some subjects may not be translatable at all. Chapter 4 will explore the limitations of datafication for the news.

One way to think about the constraints of data is by considering them as "media," for data can be characterized by their material conditions as well as their expressivity, both defining features of media forms.[27] Across the domains of science, cultural history, journalism, and real estate, data constrain how people physically and cognitively interact with the subjects of their interest, whether they are plants, books, news stories, or properties.[28]

Thus, scholars of media can aid in our understanding of how data function in the world. For instance, media theorist and designer Lev Manovich usefully juxtaposes the database with the narrative—an older system for cultural expression—by focusing on their structural differences. In fact, he argues that the database has deposed narrative to become the dominant media form of our time. "As a cultural form," explains Manovich, "database represents the world as a list of items, which it refuses to order." A narrative, meanwhile, "creates a cause-and-effect trajectory of seemingly unordered items (events)."[29] Over the course of the last half century, digital media have given rise to a variety of creative experiments, which propose how databases might function symbolically and affectively in relation to narrative. Some of those experiments, related to the themes of the book, are introduced in chapter 6.

Writing more broadly about the transformative potential of "electronic" media, theorist Marshall McLuhan, a well-known figure, if self-contradictory and controversial, might own some responsibility for the derision that the local conditions of digital media currently receive. In 1964, he wrote that "after more than a century of electric technology, we have extended our central nervous system itself in a global embrace, abolishing both space and time as far as our planet is concerned."[30] That made for a dramatic headline in the 60s. I have found that data actually preserve competing measures of space and time, however, effectively reinforcing the importance of those measures. Think of the innumerable date formats of the DPLA that haven't been so easily reconciled.

My understanding of data is plural, embedded, small, operational, and material. This is not the popular view of data, but it is in line with an emergent discourse developing among scholars at the intersection of media and information studies and STS.[31] Moreover, it is a necessary point of departure for the main subject of this book: how conceptions of data might benefit from increased attention to the local.

## DELIMITING THE LOCAL

### Local Is Not Lesser

The term *local* has become ubiquitous. It is used popularly as a modifier for widespread phenomena, objects, and actions: local food, local time, local anesthesia, or local elections.[32] In the academic world, the term has long been invoked by social scientists to

describe knowledge practices grounded in particular places, usually those inhabited by small, indigenous, marginal, or non-Western cultures.[33] Local knowledge explains how different communities make sense of things in their own way. Clifford Geertz once explained local knowledge by posing a question: "Who knows the river better, the hydrologist or the swimmer? Put that way, it clearly depends on what you mean by 'knows.'"[34] As in this example, discourses on local knowledge have often compared science, an allegedly universal knowledge system, with other ways of knowing. But such juxtapositions can echo dangerous, marginalizing descriptions of nonscientific cultures.

I use the word *local* as it has been appropriated into postcolonial and feminist science and technology studies, precisely in order to upend the dichotomy between scientific and othered cultures of knowledge production.[35] As feminist theorist Sandra Harding explains, "Postcolonial histories and studies of contemporary projects have shown that in important respects modern sciences and technologies, no less than other culture's traditions of systematic knowledge, are local knowledge systems."[36] Harding's work helps scholars to see that when modern science and technology are understood in local terms, their histories, present-day controversies, and future limitations become clearer. In recent years, STS scholars have illustrated how all knowledge systems are rooted in practices and politics related to their time and place.[37]

Knowledge systems surrounding data are no exception.[38] In STS, the term *local* is often used interchangeably with *situated*, another word that has served feminist critiques of expert knowledge.[39] Like the word *situated*, *local* is used to explain how knowledge is embodied, mediated, and historically grounded. Yet *situated* is sometimes interpreted as being about social and material conditions exclusively, while local puts more weight on the relevance of place. The concept of place, another widely theorized term, adds an awareness of spatial conditions to the investigation of knowledge practices. As geographer Tim Cresswell remarks, "Place combines the spatial with the social." Unfortunately, many practices with data are imagined to unfold in an infinitely small space. "Our consciousness of place," writes Cresswell, "all but disappears when it appears to be working well."[40] There are a number of ways that dimensions of the local, like place in general, can be suppressed from view.

At the Arnold Arboretum, many ways of knowing coexist: that of the scientists who run experiments at the institution, the arborists who tend to "their data" (the plants), the commuters who use the grounds as a way to cut across the neighborhood to the Forrest Hills train station, the foragers who appear in the spring, and even the participants in religious rituals that are known to take place on the arboretum grounds under cover of darkness. Each of these ways of knowing the arboretum is local in its own way.

## Local Is More than a Geolocation

Bringing the term *local* (and *place* by association) into conversations about how data are situated requires significant explanation. For example, space and place must be

differentiated. Space is precise, geometric, and geolocated. Place is something less easily defined: contingent on experience, defined by meaning, and susceptible to changing social designations. In his work on "critical regionalism," architectural theorist Kenneth Frampton explains that localism transcends abstract space, and instead hinges on conditions of culture and identity.[41] Moreover, locality is not merely physical. It can be virtual. In the 1980s, novelist William Gibson introduced readers to the potential of cyberspace: not only a space, but a place in which new morphologies and social forms are possible.[42] Scholars such as Michael Benedikt, Manuel Castells, and William Mitchell also theorize the meaning of place in digital life, even to the point of questioning its continued relevance.[43] Mitchell observes that "once there was a time and a place for everything; today, things are increasingly smeared across multiple sites and moments in complex and often indeterminate ways."[44] Castells portrays this as a new set of relations that transcend place. We no longer occupy spaces defined by their place-based meaning, his argument goes. Rather, we exist in a "space of flows," along which data, commodities, and capital move freely.[45] I argue that flows are still shaped by places, although in ways that have not been adequately unpacked.[46]

One of the examples in which the local transcends geolocation is the news—a topic that I address more fully in chapter 4. Local news has long been defined by a set of topics, a language, and a point of view; these dimensions of place cannot be encompassed by a pin on a map. Media and communication studies scholar Christopher Ali explains the complexity of defining local news media: "While localism can be loosely defined as the mandate for broadcasters to be responsive to their communities, localism as a symbolic category means different, often contradictory things to different people at different times."[47]

The news demonstrates just how strange yet important the local can be. Today, networks and newspapers around the United States have replaced much of what might be called "local news" with cheaper material published by national and international sources. This is a problem, for newspapers and television stations that are actually local tend to be more trusted by their audiences.[48] Meanwhile, these larger, less trusted sources are still bound by topics, languages, points of view as well as conditions such as timing, format, and systems for dissemination that shape their place in the media landscape. Indeed, the phrase *local news* can distract from the reality that all news is local, created in narrow social and historical settings for audiences prepared to receive it. The 2016 US presidential election revealed just how crucial place can be: news produced by profit-making networks based in large cities in the United States was perceived as the propaganda of urban elites, while "fake news" from Russia was blamed for unfairly skewing discussions of the election online.

On a separate note, it is useful to distinguish my focus on the locality of data from the concept of *local data*. In community-based practices, the phrase *local data* already has currency. Often understood as neighborhood scale, nonexpert, or even idiosyncratic,

*local data* can mean data made through "artisanal" practices or by some creative individual's "magic hands," rather than clearly documented and replicable processes.[49] This use of this phrase suggests results that are explicitly subjective, personal, and perhaps intentionally nonreplicable. Such forms of data may eschew the use of standards altogether and do not assume that their development is easily extensible. In that sense, this book is not about local data.

Instead, I argue that seemingly impersonal, large-scale data sets are also local. Even the US census, which purports to account for every person in the country, is entangled with local considerations.[50] It uses historically situated racial categories, only recognizes individuals with permanent addresses, and under the current administration, the 2020 census may omit "undocumented" residents who nevertheless make substantial contributions to US economic and cultural life.[51] Most of the data sets used in this book are similarly not the result of so-called artisanal practices but are nevertheless local. Acknowledging that all data are local means understanding that the phrase *local data* is redundant. As with *raw data*—now familiar as an oxymoron thanks to the work of Geoffrey Bowker and media studies scholar Lisa Gitelman, who have shown that data are always already curated or "cooked"—we should stop using the phrase *local data*.[52] Why not simply talk about all data sets in terms of their settings?

## Local Operates on Many Scales

The usefulness of the term *local* depends on its relative nature. Every local condition must be defined in its own way. As Geertz explains, "In the solar system, the earth is local; in the galaxy, the solar system is local; and in the universe, the galaxy is local. To a high-energy physicist, the particle world—or zoo—is, well, the world. It's the particle, a thread of vapor in a cloud of droplets, that's local."[53]

Similarly, in computing, *local* indicates the relative placement of a digital file: a folder is local to your hard drive, your hard drive is local in a network, and your network is local on the Internet. In common use, the term *local* gains relevance in relation to national or global contexts.[54] Despite widespread aspirations for everything to have global relevance, the local retains its importance because of the enduring stubbornness of the physical world: we cannot be everywhere at once, and every specific environment is continually changing.

Local conditions vary and operate at different scales. When it comes to housing, for example, there are numerous scales at which data face local contingencies. Houses are formally appraised in value at the scale of the individual unit. Zooming out, neighborhoods tend to rise and fall in value as a whole. At a higher level, counties create assessment models for tax purposes, which are eventually aggregated into sites like Zillow. Meanwhile, cities and states regulate home values. At the federal level, mortgages are secured, making today's inflated market possible. Housing data are shaped by conditions at all of these scales.[55]

## Local Coexists with Global

Although all data are locally made, data do not serve exclusively local needs. They must often appeal to nonlocal audiences. In fact, scholars in infrastructure studies see the local and global as complementary rather than conflicting. Susan Leigh Star, writing with information scholar Karen Ruhleder, asserts that "an infrastructure occurs when the tension between local and global is resolved."[56] For data must not only function within local structures, such as an organization, but also have a broad enough "scope" to serve relevant outside interests. Infrastructures for data are designed to work at a distance, but keeping the local and global relevant to one another requires constant coordination work. Edwards, who writes about global climate models, explains that understanding the relationship between local and global orders is a research problem: "What made it possible to see local forces as elements of a planetary order, and the planetary order as directly relevant to the tiny scale of ordinary, individual human lives?"[57]

At the Arnold Arboretum, data may be shaped by local conditions, yet they serve a combination of needs, near and far. While accessions data help staff to manage the collection on a day-to-day basis, they are also relied on by researchers from institutions around the globe who collaborate with local arboretum staff as well as by tourists who visit from abroad and delight in using data tagged directly on the plants to navigate the otherwise-exotic collection.

Nevertheless, there is no global experience of data, only an expanding variety of local encounters. Data travel widely, but wherever they go, that's where data are. For even when data escape their origins, they are always encountered within other significant local settings.[58]

## Local Is Not an Unquestionable Good

The term *local* is a point of pride in advertising. Local ownership, sourcing, and artisanship are all reasons that one should, purportedly, choose to "buy local." Moreover, as recent criticism of globalization gains momentum, local can come to be seen by some as unequivocally positive. But local can also mean exclusionary, narrow, or even oppressive. In recent years, right-wing nationalist movements in the United States have sometimes aligned themselves with localism. Many examples in this book reveal how local conditions can create problematic data or data practices. The DPLA offers opportunities to see how data created to manage collections of cultural history can reify racist social categories that are local to places in the United States. Why is it that curators at the Smithsonian, for instance, hold data on the race of "black" and "African American" artists, but not those of "whites"? This absence paradoxically legitimizes white supremacy by suggesting that whiteness is normal, not in need of special attention, and the default from which all other racial categories must be distinguished.

Returning to the example of the news, we now know that choosing to see only what is local can lead to the emergence of filter bubbles—a phenomenon that online

organizer Eli Pariser explains can arise when data-driven algorithms return only the results that we expect.[59] All other perspectives are "filtered" out. Filter bubbles have been blamed in recent years for the extreme polarization of US politics, leading Democrats and Republicans alike to dismiss the other side as uninformed, or even misinformed by what they deem to be "fake news." As we learn to engage with the local conditions of data, we must also become aware of when those conditions need to be protected and when they are more justifiably challenged.

### Local Often Means Invisible

A local perspective is not easy to maintain, particularly when interfaces to data ask us to accept those data out of context. Appreciating that all data are local requires effort. Unfortunately, minding and maintaining the context of data is frequently invisible labor, best understood as a kind of care.[60] Feminist scholar and political theorist Joan Tronto, writing alongside Berenice Fisher, defines care as "everything we do to maintain, continue, and repair 'our world' so that we can live in it as well as possible. That world includes our bodies, ourselves, and our environment, all that we seek to interweave in a complex, life sustaining web."[61]

Seeing data work as care is in keeping with Harding's feminist and postcolonial focus on the local. Librarians and archivists explicitly engage in care through data. In chapter 3, I write about conflicts between librarians and internet scholars in the formation of the DPLA. Librarians "deal in privileges," one of the DPLA staff warns, adding, "They keep data away from the public."[62] Another cautions, "Librarians can be overly concerned with the value of their holdings and themselves."[63]

Such comments are disparaging of important care work that is necessary for the creation, operation, and continued maintenance of data in every setting. We must all do more to actively care for our data and any vulnerable subjects that they represent. When such work is derided or undervalued, it perpetuates a long history of degrading care.[64] Seeing data as local necessitates acknowledging care for data as well as those who carry out that work. Especially for those readers who identify as men, as I do, adopting a feminist perspective can be a crucial step toward challenging gendered notions of objectivity applied to data. Treating data subjects abstractly, avoiding intimate relationships with those subjects, and holding them at a comfortable distance does not make one more rigorous but rather less well informed.[65]

### CONCLUSION

In practice, accepting that all data are local means engaging with data settings instead of simply data sets. Today, data are too often harnessed as discrete tools to enable analytic work at a distance. The perception of data as immutable yet mobile sustains this view. But we don't have to distance ourselves to make use of data. Indeed, data have been used in many instances—by scientists and curators, among others—as a

means of keeping subjects close across long stretches of space and time. Local practices necessitate forming close relationships with not only data but the conditions in which those data are manifest too. Thus, this book is meant to not only reflect on the representational and rhetorical aspects of data from a distance—what values and assumptions they embody—but also engage directly with the knowledge systems that data construct and maintain. What does it mean to take a local stance in data gathering and analysis? How can we get local audiences to care about data? What do models of local practice with data look like? These are some of the questions implicitly posed throughout the book.

In the next chapter and first collections data case, I examine the various roles that place can play, using accessions data from Harvard University's Arnold Arboretum, one of the world's largest collections of trees, vines, and shrubs. Its story illustrates complex ways in which data can be about, in, or from a place as well as how the profile of a place might be understood in terms of its data.

# 2