




All the Homes: Zillow and the Operational Context of Data

Yanni Loukissas^(✉) 

Georgia Institute of Technology, Atlanta, GA 30308, USA
yanni.loukissas@lmc.gatech.edu

Abstract. Zillow, an online real estate marketplace that seeks to make information available about “all the homes” in the United States, tells us that “data want to be free”. But a close analysis reveals that Zillow works to ground data: to put data into an operational context. I use the phrase “operational context” to denote a setting in which data—for real estate: current listings, tax assessments, and other digital property records—are meant to be fully understood. This paper examines the design of operational contexts for data as well as their cultural and political significance, using Zillow as a case. Zillow was founded in 2006, at the height of the housing bubble. Although practices with real estate have been under scrutiny ever since, the treatment of real estate data has not. This paper examines how Zillow operationalizes data for the housing market through a combination of analytical, discursive, and algorithmic devices. These dimensions of operational context are less about establishing the truth of data than a level of tractability for prospective buyers and sellers. The operational context for data is not derived from a neutral retrospective view (i.e. where the data come from). Rather, it is a matter of connecting data to an existing cultural system, defined by inherited practices, concepts and affordances that support specific use cases. Operational context can enable interpretation and action based on data, but it can also reify the power of a dominant culture.

Keywords: Data · Context · Housing

1 Introduction

Whether you are looking for a place to live, a good meal, upcoming events in your area, or a ride to work, a new economy of apps stands ready to serve, through interactions that can be carried out on any networked computing device. The data that enable these transactions are created at the local level, collected by civic institutions or crowd-sourced from the users themselves. However, beginning in the United States, they have been rapidly mobilized by data brokers [1, 2], who build and maintain national- or international-scale data infrastructures for profit. The boosters of this new “smart” lifestyle are ushering in a new kind of individualism tailored for affluent and tech-savvy urban dwellers. Yelp, an online directory of restaurants, shopping, and personal services, can make sure you “connect with great local businesses” [3]. Nextdoor, a place-based social media platform, invites you to “discover your neighborhood” [4]. Uber, a networked car service, equates “getting there” with personal

freedom: “your day belongs to you” [5]. Zillow, the real estate website, will help you “find your way home” [6]. These information systems promise not only access to data, but also the *operational context* to act on them.

By operational context, I mean the analytical, discursive, and algorithmic systems that connect abstract streams of zeros and ones—the bare-bones definition of contemporary digital data—to the concepts and resources that support their effective use. Analytical elements, such as a map, graph, or timeline, help users see meaningful patterns in data. Discursive elements offer you ready-made narratives with which to frame these data. Algorithmic elements enrich data, by generating new value from existing inputs. I call all these elements of context *operational* because they transform data from simple representations of the city—in terms of prices, distances, and rankings, for instance—into drivers for local and highly personalized behavior. Data, as well as the proper context to use them, are increasingly desirable for affluent consumers in cities across the United States and in many places abroad.

In this paper, I bring an interpretive approach to the question: what does it mean to put data in context? This entails a cultural analysis of the interfaces, discourses and processes that shape information systems, such as those described above. The question of context is of deep relevance for those who study and design systems that mediate relationships between people and data. Although the term *context* is widely used in both academic and popular writing, its relationship to data is not well understood [7]. I argue that scholars and designers of information should see context in operational terms. On the face of it, connecting data to an operational context is a pragmatic problem of supporting data use. However, establishing operational context is also a problem with important social and even political consequences: who can use data? how they can use data? and what they can use data for? Indeed, operational contexts establish the subject positions that users of data are expected to adopt. In the case of Zillow, which will be my main focus throughout the paper, you can be a prospective home buyer, renter, or seller; or you can be a relator. No homeowner or resident, according to Zillow, is outside of the market. As I will show, the potential operational contexts for data are always local and always multiple; they enable different forms of engagement and interpretation, with implications for what data appear to say and who they are made to speak to.

In order to elucidate this point, I examine the concrete problem of putting housing data into operational context. The values of homes in the United States, and other countries where property is on the market, have long been determined, in large part, by context. The perceived worth of a home is not determined solely based on its age, square footage, or the number of bedrooms and bathrooms it contains. Home values fluctuate based on comparable sales in the area, changes in the neighborhood itself, interest rates, and even the time of year. What counts as context when it comes to pricing a home? The seller and the buyer are the ultimate arbitrators of that. However, professionals—including realtors, lenders, researchers, developers, and, more recently, information technologists and designers—seek to influence perceptions of context in housing.

Today, context in housing is increasingly assessed through data. But although the housing crisis of 2007 raised important questions about the way we finance housing in the United States, it has not raised parallel, and necessary, questions about the way we

use housing data. I intend to address some of those questions here, asking: how are housing data put in context, and how does this context then shape perception and action in public life? Zillow, a prominent example of information design applied to housing data, takes input from public and private sources, such as tax assessments and sales records, in nearly every municipality in the United States. It has used these resources to shape—as much as any other information system—the context in which non-experts understand housing through a combination of analytical, discursive, and algorithmic choices. Zillow demonstrates a range of ways for putting data in operational context and the stakes in doing so. In fact, the frames through which we examine housing data—specifically the operational context as defined herein—impact access to affordable housing on the market. For this reason, it is necessary to reconsider the settings through which we look at, talk about, and calculate with housing data. However, before I delve into the specific elements of operational context cultivated by Zillow, I would like to distinguish operational context from other models of context used in the study and design of information systems.

2 Data in Context

My operational perspective on context differs substantially from the dominant modes of accounting for context in information systems. Theorist Paul Dourish sums up the prevailing views in his article “What We Talk about When We Talk about Context” [8]. He specifically addresses definitions of context in ubiquitous computing, an area of research that explores the potential for computers to be distributed throughout the range of human environments. Dourish juxtaposes a “representational” model of context, pursued by the majority of researchers in computing, with an “interactional” model, grounded in phenomenological inquiry—an area of philosophical thought focused on understanding individual human experience.

Dourish explains that in the representational model, context “consists of a set of features of the environment surrounding generic activities” [8]. Representational context is easily delineable, stable, and separable from the subject itself. Meanwhile, in an interactional model, writes Dourish, “context isn’t something that describes a setting. It is something that people do” [8]. As such, the context of any event or object can vary enormously depending on whom you talk to and when. Interactional context is relative, dynamic, spontaneous, and arising from activity. However, in order to avoid further confusion over the term, Dourish suggests that we leave aside the notion of context altogether. Instead, why not think about “practices” as the forming the settings for human interactions with computers?

However, the term context is not disappearing from use. As we seek more generalized solutions for managing human relationships with data, the concept will need further refinement. Representational context is certainly inadequate. But neither can context be summarized as something determined spontaneously, in an unselfconscious moment—what Dourish terms interactional context. Rather, I argue, context should be understood as a “cultural system” [9], composed of inherited practices, concepts, and affordances. Although an operational context for data does not determine the way data are used, it provides a setting that shapes the roles and forms of reasoning adopted by users.

The operational context of data is the setting in which participants are equipped with the resources and subject roles necessary to access, interpret, and take action on predetermined objects of attention. However, to put it that way is to suggest that operational context is something settled and uncontested. That is not the case. Contexts that operationalize data are always under construction. Moreover, disputes over operational context are common, sometimes with striking significance, as the case of Zillow reveals.

In a domain like housing, operational context can have the highest stakes. Thus, we must ask, what does operationalizing data enable? For operational context is not just an issue of knowledge, but of use. Scholars and designers of information systems should consider how operational contexts are rooted in normative cultural assumptions about what data can and should do.

Zillow operationalizes data through a context that combines analytical, discursive, and algorithmic elements. The analytical setting of Zillow is defined by the functionality of its map. Placing data on a map enables comparative reasoning, but only about things that have a geographic dimension. Zillow's main discursive setting is that of "public data." Data in the public realm are increasing accessible, but at what cost? Finally, Zillow uses an automated valuation model, the "Zestimate," to contextualize data. This algorithmic setting offers an interpretation of data, but through a set of opaque and speculative rules. These three dimensions of operational context are less about establishing the (capital "T") truth of data on property values than creating traction with a community of users. In the case of Zillow, the users are expected to be buyers, renters, or sellers engaged in the housing market. Kitchin has used the term "assemblage" to describe the many facets of such interactions with data [1]. But operational context is much more precise and structured than the term assemblage suggests.

3 The Case of Zillow

Zillow is a leading online real estate marketplace seeking to redefine the context in which we understand housing through access to and analysis of data. The name is a portmanteau created by combining the words zillion and pillow (where you rest your head). The company was founded in 2006 by Rascoff and Humphries with the goal of estimating the value of every home in the United States. It is not a licensed real estate firm, which would require that the company submit to licensing rules and regulations in every state that Zillow practices. However, it has strategically intervened into the real estate market in a way that has changed the work of realtors and other professionals in the industry. Zillow is not the first web company to tread into real estate. It is just one of the many data brokers that seek to produce surplus value from available data on housing [10]. Moreover, Zillow's way of operationalizing data is not original or unique. However, its recent purchase of Trulia—another major platform for home listings focused more on user experience than analytics—has consolidated Zillow's position as the market leader in the United States.

"Data want to be free," explains a representative of Zillow during a routine webinar. Setting data free sounds like a laudable, emancipatory goal. However, much

of what Zillow does is to ground data in existing systems: analytical, discursive, and algorithmic. Indeed, Zillow offers important lessons on how to put data into operational context. But their approach should also give us pause, for they demonstrate that operational contexts are not neutral. They make data available to specific groups and support targeted actions. I will discuss three elements of Zillow’s operational context. Each of these elements establishes relations that are important for working with data: access, interpretation, and action. These strategies for operationalizing data were not invented by Zillow. They are cultural forms that exist independently of the platform. However, Zillow assembles its operational context in a way that has significant implications for the way we understand housing data. For Zillow is invested in furthering the consumer culture of property, by creating a seemingly classical economic setting in which individuals are given access to information and encouraged to make rational choices based on their own self-interests. But the effects of this setting are damaging in ways that Zillow obscures. For although users may believe they are independent actors, the demand they place on the market works to increase the value of all property in an area, and limit the availability of affordable options. Zillow not only supports this system, but increases anxiety about its instability through the introduction of its Zestimate algorithm. In the sections that follow, I will unpack all three dimensions of Zillow’s operational context, by explaining how they are constructed and by reflecting on their effects on data and housing.

3.1 Analytics

For Zillow, putting data in context starts with positioning it on a map. Zillow data—listings of properties for sale, for rent, or otherwise of interest (i.e., foreclosure or a category simply labelled “make me sell”)—appear as colored dots outlined in white and placed on a grey background. The uniform setting for listings data is marked only by a faint network of streets, parks, bodies of water, and place names. Hovering over a dot brings up a price tag, including the number of bedrooms and bathrooms, the square footage, and a small thumbnail image. This intentionally generic setting—the same everywhere across the geography where Zillow lists properties—frames our understanding of housing data, not by showing the conditions of their production (representational context) or how they might have been used in the past (interactional context), but by suggesting what can be done with them today. The map is analytical and operational.

A venerable technology for visual reasoning, maps are recognizable and accessible to most of Zillow’s users [11]. They offer a structure for making sense of data through spatial patterns; they show where the listings are located. This enables comparative readings of listings (i.e., “these listings are close to one another”) as well as readings of each data point within a matrix of surrounding features (i.e., “these listings are close to a park”). These are operational relationships—they can serve as the basis for consumer decisions about real estate, a domain in which it is said that the three most important indicators of value are location, location, and location [12]. But the map does not merely register the locations of real estate in the real world [13]. Rather, the map produces a reading of location using a narrow set of visible relationships (i.e., to select streets, bucolic parks, and highly ranked schools). In this way, the map participates in

the production of reality for real estate, by establishing or confirming conceptions about what conditions of location determine value [13, 14]. Thus, putting data in the operational context of the map is not a retrospective practice. The map does not reunite data with some preexisting setting. Zillow's map is operational because it stimulates actionable interpretations of location and its relationship to home value.

But maps do not “unfold” in isolation [15]. Zillow's map is framed by other media and modes of access to the underlying data. Above the map is a search bar, with filters for listing type, price, number of beds, and more that can be applied to further narrow the number of listings displayed. To the right of the map is a column of property images, mostly facades. Each is annotated with further details about an individual listing such as the number of days on Zillow, the name of the listing agent, and the type of sale (house for sale, pre-foreclosure, lot/land for sale). These images in turn can act as links to a full-screen view of an individual listing.

The additional elements of the Zillow interface serve to put the map itself in relief. They help users interpret the map as a collection of commodities: locations valued because of their potential to be bought and sold, not because of their historical significance as places or the significance of the people who live there. Thus, the analytical elements of context illuminate a number of things: which data points matter, the relationship between the points, the meaning of the space in between them, and the connection between data and any secondary media.

3.2 Discourse

Beyond the setting established by Zillow's map interface, the company offers a discursive component of operational context. It serves to establish and stabilize Zillow's use of data as a legitimate representation of the world of real estate. Among the most important discourses that Zillow invokes is that of *public data*. In a list of frequently asked questions on their website, Zillow explains *public information* (a related but more expansive term than *public data*) as the way and the reason it knows about your house:

Zillow receives information about property sales from the municipal office responsible for recording real estate transactions in your area. The information we provide is public information, gathered from county records [6].

The term *public* tells us that Zillow's map is based on data from an open and authoritative source: the municipal office in “your area.” This legitimizes both the data and Zillow's use of them. For although Zillow operates outside of the boundaries of any particular municipality (remember it is not licensed anywhere), it is making fair use of data created by and for the people. Moreover, invoking the public context of their data protects Zillow from requests by homeowners to have it removed from the site. Their reasoning: it is not private data. And although Zillow acknowledges that municipally created data may contain errors, the company takes no responsibility for them. After all, the data are not created by Zillow. It is up to homeowners to show proof of errors that might affect public perceptions of the value of their house. Ironically, all these assertions about Zillow's rights with respect to the data are wrapped in the language of public empowerment:

Our mission is to empower consumers with information and tools to make smart decisions about homes, real estate, and mortgages. For this reason, we do not remove public record property data from Zillow unless it is shown to be erroneous [6].

Beyond their use of public sources, Zillow also cultivates a perception of their map as a virtual public space in which data from private sources might be made broadly accessible. There are two ways this can happen: realtors can contribute their own listings—and pay a fee to have their profiles promoted in association with those listings—or owners can contribute “house facts” in order to improve the online image of their property. A former employee of Trulia explains, “if you are able to give people a real-time value of their home, they are going to check that value and ask: what can we do to update that value? The Zestimate is a powerful consumer engagement instrument” [16]. By making their database open to public reading as well as public writing, Zillow fashions itself as the “Wikipedia of housing”: a democratic, free, and transparent context for sharing data publically [16].

In tension with the discourse of public data is that of the personal journey. “Find your way home” is the welcoming message on the Zillow front page. “You are in the driver’s seat,” we are told. The implication is that Zillow is a powerful vehicle that we can use in our travel towards homeownership. This second discourse positions the platform as a navigational aid in an individualized search for *home*—a term used as a synecdoche for personal comfort, security, and belonging—through a bewildering landscape of consumer options. While the space of data on Zillow is public, the journey through that space is private and the implication is that it should be guided by individualized interests, as opposed to the public good. The result is the creation of a public space in which everyone has access, but no one is equal. Everyone comes to the map with different resources for buying or selling and Zillow lets you know right where that places you. Thus, the discursive elements of context define the relationships between people and data: who owns it, manages it, or uses it, and who doesn’t? and what stories about the nature of the data justify these attachments or exclusions?

The map, its media annotations, and an overarching discursive framing put Zillow’s data in context, a context that isn’t a reconstruction of the origins of data, but rather an operational setting that makes the data actionable. But the context marshaled by Zillow extends beyond these analytical and discursive elements. Zillow has been successful in large part because of an additional computational layer of operational context that it brings to the existing set of housing listings.

3.3 Algorithms

Zillow’s “rules of real estate” establish the final dimension of operational context I will discuss [17]. The company not only accumulates data from a variety of sources but also extracts surplus value from those data—in the form of computationally generated predictions. Using the data that Zillow has assembled on sales and historic valuations of homes in a particular area, the company is able to produce estimated values for properties that are not currently on the market. This process of triangulating property values is called an automated valuation model. The results of Zillow’s model, comically called “Zestimates,” can be applied to nearly every home in the United States. Today, in 2017, a Zestimate is calculated for about 100 million homes nationwide

using public data as well as data contributed by relators or homeowners. The physical characteristics of a home, such as its location, square footage, number of bedrooms and bathrooms, as well as past sale prices of the home and comparable homes nearby, are analyzed using proprietary valuation rules. Instead of relying on a single complex model of the entire United States market, Zillow relies on simpler, localized models (sometimes at the scale of a single street) to account for different market situations [17]. Moreover, their models are dynamic. Home values are discarded every night and built again in the morning using fresh data that incorporate changing conditions and new information from local sources.

Highlighting the operational context they bring to existing data, Rascoff and Humphries, the founders of Zillow, argue that it is not their data, but the Zestimates they extract from those data that differentiates them from other online real estate sites. Indeed, Zillow claims it has a pulse on market conditions in local areas across the country, and, moreover, that it can make accurate predictions on where the market is headed in the near future. For Zillow, the future is just another discursive element to be marshaled in operationalizing housing data.

Such algorithmic elements of operational context are a form of what Janet Murray and other digital media scholars call procedurality, “the computer’s defining ability to execute a series of rules” [18]. More specifically, the Zestimate might count as a form of what Ian Bogost terms “procedural rhetoric” [19]. To adapt this framework, we might say that the Zestimate makes claims about how property values operate. It claims that not only square footage or the number of bathrooms and bedrooms, but a host of other, sometimes specifically local characteristics are intertwined with value. But we shouldn’t see this as simply a representational system, as in the case of the persuasive games that Bogost studies. Rather, Zillow is an operational part of the way that the housing market works today. Zillow’s rules form a system that homeowners can interact with by updating their home facts or simply checking their Zestimate regularly.

The Zestimate is an “algorithmic system,” an arrangement code and people [20]. Understanding such systems means examining not only the technical details, but also the cultural practices of its creators and users. Such efforts are further complicated by the fact that the Zestimate, like many commercial algorithms, is a closely guarded secret and requires insider knowledge to be decoded. However, we can learn much from simply looking at how the Zestimate is discursively framed.

Like the map, the Zestimate has its own discursive elements. Buyers and sellers are encouraged to put the Zestimate itself into context using feelings, facts, comps, pop psychology, and the general economic climate. After all, Zestimates may be wrong; the company acknowledges that they are within 5% of the actual value (sale price) of the home just half of the time. Ultimately it may not matter, if they lure you into a conversation with realtors and other professionals who pay Zillow to stay in business.

Indeed, Zillow doesn’t depend on the accuracy of its Zestimates to make money. It thrives on subscriptions from realtors and other related professionals. This is an advertising model of revenue, not unlike those used by other web platforms that host social media, such as Facebook and Twitter. From an operational perspective, the algorithm is good enough if it catalyzes a certain kind of social relation between subscribing real estate professions and potential clients. Zillow is concerned less about the truth of data than its tractability: the holding power it exerts on users.

4 Conclusion

The purpose of this paper has been to make the possible contexts for data more visible. But this requires a clear understanding of what context can mean. The case of Zillow demonstrates that the context for data need not be simply representational: an account of the setting in which the data were made. Nor is context exclusively interactional: the spontaneous result of an engagement with data. Rather, putting data in context can be operational: a matter of connecting data to an existing cultural system, defined by a combination of practices, concepts, and affordances that are meant to support data's use. Through analytical, discursive, and algorithmic devices, Zillow has constructed a captivating operational context that supports the use of data to buy, rent, or sell housing. But the context that Zillow has assembled also reifies dominant and deeply problematic relationships inherent to our market-based culture of property.

While technologists and designers at Zillow are working to establish operational contexts in which consumers might make the best personal choices based on housing data, others, community organizers for instance, may seek to reveal another operational context: one that calls the market into question. For example, by: 1. using a timeline as the main analytical setting, to illuminate the chaotic history of market values; 2. invoking the discourse of gentrification, to call attention to the damage caused by rapid market change; and 3. calling for regulatory rules rather than only the rules of the market to set home values. Conflicting contexts enable different ways of imagining and enacting the future of housing through data. They both implicitly accept that data are now a necessary tool for addressing the problem of housing, which has reached a scale that would be difficult to contemplate otherwise. Indeed, the operational context of housing data has become a site of contestation, which will determine how housing evolves and for whose benefit.

Although this paper is illustrated by examples from the domain of housing, where the stakes for accessibility, interpretation, and action are high, there are many other domains from which important examples of operational context might be drawn: health, crime, and climate change, to name a few. In any of these areas, practitioners who design for information do not act autonomously. Rather, they must connect with existing communities of use to support data-based action. For contexts are not merely representative or interactional; they are deeply rooted in operational cultural systems.

References

1. Kitchin, R., Lauriault, T.P.: *Towards critical data studies: charting and unpacking data assemblages and their work*. Social Science Research Network, Rochester (2014)
2. Small, B.: *FTC report examines data brokers*. Federal Trade Commission (2014)
3. Yelp Homepage. <https://www.yelp.com>. Accessed 15 Dec 2017
4. Nextdoor Homepage. <https://nextdoor.com>. Accessed 15 Dec 2017
5. Uber Homepage. <https://uber.com>. Accessed 15 Dec 2017
6. Zillow Homepage. <https://www.zillow.com>. Accessed 15 Dec 2017
7. Seaver, N.: The nice thing about context is that everyone has it. *Media Cult. Soc.* **37**(7), 1101–1109 (2015)

8. Dourish, P.: What we talk about when we talk about context. *Pers. Ubiquit. Comput.* **8**(1), 11–30 (2004)
9. Geertz, C.: *The Interpretation of Cultures: Selected Essays*. Basic Books, New York (1973)
10. Federal Trade Commission: *Data Brokers: A Call for Transparency and Accountability*. Government of the United States (2014)
11. Crampton, J.W.: *Mapping: A Critical Introduction to Cartography and GIS*. Wiley, Hoboken (2011)
12. Pearson, D.: Location, location, location. *The New York Times* (2009). <http://www.nytimes.com/2009/06/28/magazine/28FOB-onlanguage-t.html>. Accessed 13 Nov 2017
13. Robinson, A.H.: *The Look of Maps*. University of Wisconsin Press, Madison (1952)
14. Baudrillard, J.: *Simulations and Simulacra*. Sheila Glaser (trans). University of Michigan Press, Ann Arbor (1984)
15. Kitchin, R., Gleeson, J., Dodge, M.: Unfolding mapping practices: a new epistemology for cartography. *Trans. Inst. Br. Geogr.* **3**(3), 480–496 (2013)
16. From an interview by the author (2016)
17. Rascoff, S., Humphries, S.: *Zillow Talk: Rewriting the Rules of Real Estate*. Grand Central Publishing, New York (2015)
18. Murray, J.: *Hamlet on the Holodeck*. MIT Press, Cambridge (1998)
19. Bogost, I.: *Persuasive Games: The Expressive Power of Videogames*. MIT Press, Cambridge (2007)
20. Seaver, N.: Algorithms as culture: some tactics for the ethnography of algorithmic systems. *Big Data Soc.* **4**(2) (2017)