Nicholas Bolten

Equitable Network Modeling of Diverse Modes of Built Environment Pedestrian Navigation

Nicholas Bolten

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2020

Reading Committee:

Eric Klavins, Chair

Anat Caspi

Sreeram Kannan

Program Authorized to Offer Degree:

Electrical and Computer Engineering

University of Washington

Abstract

Equitable Network Modeling of Diverse Modes of Built Environment Pedestrian Navigation

Nicholas Bolten

Chair of the Supervisory Committee:
Professor Eric Klavins
Electrical and Computer Engineering

Pedestrian informational needs are currently not well-met by existing data infrastructure or public-facing applications, limiting the public's ability to reliably plan trips through public spaces and to understand existing inequities in infrastructure such as network gaps that prevent access to wheelchair users. Modeling pedestrian accessibility requires addressing multiple key problem domains simultaneously, including applying techniques borrowed from the modeling of non-pedestrian transportation networks as well as bringing in outside and new techniques. This dissertation will lay out the numerous problem domains in creating a truly adequate pedestrian network data model, populating that data model, and leveraging it to provide appropriate information to individuals and analysis for use by agencies or the public. The process of building and studying the AccessMap website, an interactive, personalized accessibility map is presented to illustrate the technical and informational barriers that currently stymie work on this issue, including the lack of a standard data schema for pedestrian networks, the need to make use of non-standard data formats, the need to collect fundamental pedestrian network information, the definition of appropriate cost functions for pedestrian mobility, the lack of information regarding pedestrian informational requirements and preferences, and challenges in communicating pedestrian preferences. This work also discusses graph analysis and advocacy enabled by a city-scale pedestrian network and appropriate pedestrian cost functions for graph traversal.

TABLE OF CONTENTS

Table of Contents

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First, I would like to thank my advisors Dr. Anat Caspi and Prof. Eric Klavins for their support of my PhD research and mentorship. Dr. Caspi inspired and continually supported this thesis work, taking chances on an under-researched field and a graduate student looking for a new direction. I cannot thank Dr. Caspi enough for her support of myself and this work. Without the support and initial guidance of Prof. Klavins, I would never have started down the path to PhD research and I will always be thankful for the opportunities he provided.

Aside from my advisors, I would like to thank the rest of my committee, Prof. Sreeram Kannan, for his insightful words and encouragement, as well as for his work teaching a course that inspired much of the graph analytic work discussed in this thesis.

Finally, I would like to thank the family and friends that have supported me and my PhD work throughout the last many years. In particular, I would like to thank my partner, Dr. Sundipta Rao, and my family in Montana for their constant emotional (and even technical) support and encouragement.

CHAPTER 1. INTRODUCTION

1.1 THE CHALLENGE OF PEDESTRIAN NAVIGATION

Ubiquitous features of the pedestrian built environment pose severe outdoor barriers for individuals with limited mobility, as demonstrated by numerous studies [5,34,35,37,41]. Such features include sidewalk surface type, evenness of terrain, stairs, curbs, elevation changes and construction. While some criteria, such as elevation or sidewalk location, change infrequently, others are transient, such as construction or weather conditions. Further, in a formative study, Hara et al. [34] determined the mobility impact of a given barrier depended on the participants' level of mobility and the type of the assistive device they used.

For example, while grassy surfaces or curbs can be negotiated by cane users, they pose a significant barrier to wheelchair users, who must seek detours. Even among wheelchair users, those with motorized wheelchairs can negotiate surfaces or slopes that are impossible for manual wheelchair users [5]. Diverse factors - like strength and weight of the user and their assistive device, familiarity with the assistive device, and personal comfort - are all considerations when determining the most appropriate pedestrian path for a given individual. In addition, some features of the built environment, like curb ramps that encompass an entire sidewalk corner, may be desirable for many wheelchair users but pose a challenge to those with visual impairments who need cues about the direction in which to cross the street.

Figure 1.1: **Common features of the urban environment determine pedestrian accessibility, but they are not accounted for in trip planning services.**

The labeled photograph highlights a subset of the informational gap faced by pedestrians, particularly those with limited mobility. Knowledge of the location of sidewalks, crossings, curb ramps, and construction is necessary to provide a safe, navigable route. The map below the photograph displays all segments traveled in 1000 routes from the Google Directions™ service, starting at the blue dot in Seattle, WA. A wheelchair ramp with an incline greater than 8.33% exceeds ADA requirements. Like most trip planning services, Google Maps™ does not take elevation into account for pedestrians, and more than 1 in 3 routes in this example included at least one segment with a greater than 8.33% incline.

Heterogeneous pedestrian information requirements imply the need for detailed, specific data on the built environment rather than broadly categorizing descriptors, such as “wheelchair accessible.” While geodata for automobile road networks are widely available in the United States, digital sidewalk geodata are scarce, have no standard format, and, if available, are produced and maintained by local municipalities, not coordinated at the regional level [39]. Existing trip planning services cannot leverage pedestrian information that is not available; therefore, they often produce routes that include potential mobility barriers (Figure 1.1).

Kasemsuppakorn et al. [42] studied the routing preferences of 5 participants with mobility impairments. Although subjects had to travel an average of 14.64% farther, they rated personalized routes as having better path quality than shortest-distance ones[42]. While this study demonstrated the benefits of personalized routing, it was necessarily limited in scale, as pedestrian trips had to be planned by hand. Scaling up studies on personalized routing requires automation of planned pedestrian trips, which is another challenge exacerbated by a lack of pedestrian data.

1.2 REQUIREMENTS OF AUTOMATED (PEDESTRIAN) TRIP PLANNERS

Typically, streets are modeled as centerlines. This semantic decision follows from the fact that car traffic must stay in lanes/roads, and is concerned with reaching destinations based on connections among roads. By contrast, pedestrians can more freely navigate spaces without lanes or an expectation of standardized environmental features. Thus, this centerline model does not translate well to pedestrian paths, which must often be navigated more like 2-dimensional spaces like plazas or sidewalks for which half of the path is blocked by street furniture like lamp posts, sandwich boards, or trash cans.

Automatic route planning uses optimization algorithms that operate on connected graphs, where each edge represents a segment of a transportation network and is assigned a cost (or weight) based on the difficulty of travel. Costs are typically determined by a function that takes into account factors contributing to a path's difficulty, and the optimal path between any two vertices in a connected graph has the lowest summed cost. For a transportation routing graph to be useful, it must (1) have nearly complete edge coverage, so that no essential paths are missing, (2) be spatially accurate so that graph vertices can be generated where paths meet, and (3) data relevant to calculating edge costs must be relational such that they can be associated with a specific path (e.g., speed limits for streets). For automotive routing, street centerlines constitute graph edges and street intersections constitute graph vertices. Street data are ubiquitous, curated at the national level in many countries, are of direct interest to companies in industries such as the transportation of goods, and are the focus of transportation network analysis. Thus, they are near-complete and spatially accurate. Relational data for calculating edge costs need only be related to streets spatially or by metadata, and are most often in the form of travel time estimates for a small number of travel modes (driving a car and public transportation) using information like speed limits or wait times from aggregated GPS data.

In contrast, pedestrian path geodata are curated by municipalities primarily for inventory and liability purposes and are published less frequently than street geodata. Because the purpose of curating data is not focused on treating pedestrian ways as a transportation network, when pedestrian path geodata are available, they are stored in a variety of incompatible structures, including the metadata of streets ("a sidewalk is on the left side of the street line"), separate polylines, and disjoint polygons derived from aerial orthoimagery. Recent work has shown that actual sidewalk adjacency is an important factor that pedestrians consider when choosing a walking route[26], but sidewalk data can contain frequent breakages or inaccuracies that prevent the creation of a routing graph. Furthermore, the sidewalks must be connected by supplemental paths such as crosswalks, bridges, and separate footpaths to have a complete pedestrian graph.

Compounding this challenge, a pedestrian routing graph also requires a cost function by which to score routes on an individual basis. The heterogeneity of potential barriers impacting accessibility highlights the need for a cost function that determines the best balance between several mobility criteria, and is broadly discussed in the literature[5,34,42,63]. However, the data needed for a pedestrian routing cost function, such as curb ramps, construction information, and steepness, are often curated separately from sidewalk datasets, and require significant post-processing to associate with pedestrian paths, a problem amplified by spatial inaccuracy in sidewalk locations. Making pedestrian wayfinding widely available can therefore be posed as a data integration problem: data and data structures must be standardized to support a pedestrian transportation network, the spatial accuracy of existing datasets must be improved and related to the pedestrian transportation network, and the many stakeholders involved in producing and consuming pedestrian geodata, including municipal agencies, must have the means and tooling to integrate their data resources.

In order to draw from the successful paradigms of car-centric navigation, we chose to work around these challenges by creating custom cost functions for routing and drawing more pedestrian paths than are typically collected: in addition to sidewalks, street crossings, and paths through parks, we added pedestrian paths that may be less obviously visible, such as the path that extends from a sidewalk centerline, through a curb ramp, and across part of the street to get to the crosswalk, or primary paths through pedestrian plazas. Eventually, it may be more appropriate to collect pedestrian data as truly 2-dimensional features, but this will require interpretation into connected paths in order to exploit efficient routing algorithms.

1.3 Pedestrian accessibility maps and previous work

Table 1.1: Comparison of known interactive pedestrian map technologies

	Primary pedestrian paths	Crossings	Curb ramps	Stairs	POI accessibility	Accessible path routing	Individualization	Transit trip planning	Street level accessibility visualization	City-scale+
AccessMap[76]	■	■	■			■	■			■
OpenRouteService[77]	■	■				■				■
PAM[41]	■	■				■	■			
Hashemi & Karimi[35]	■	■				■	■			
WheelMap[48]					■					
AXS Map[78]					■					
Soundscape[79]	■	■		■						
PathVu[80]	■	■	■							
Google Maps™								■	■	■
Apple Maps™								■		■
Bing Maps™								■		■

There are a wide array of concerns that must be covered to build a general-purpose accessibility map, and past work in addressing them are shown in Table 1.1. AccessMap, the work driving this dissertation, operates a personalizable accessibility map at city scale with user-defined cost functions and a wide set of pedestrian paths, but lacks multi-modal transit functionality, stairs, and POI accessibility. Karimi et al[41] gathered pedestrian network data and implemented cost functions and automatic trip planning, but operated at a limited scale (a single university campus) and no longer has a functioning interactive map. Hashemi & Karimi[35] identified trade-offs that inform the definition of personalized cost functions, but did not create a public-facing map product. WheelMap[48] and AXS Map[78] crowdsource accessibility labels for public POIs like businesses or other venues and have successfully gathered a large number of labels in OpenStreetMap, but are limited in scope: the accessibility of a venue is identified but not the paths to it. In addition, their approaches take a one-size-fits-all view of accessibility that does not account for pedestrian diversity, requiring third parties to classify the accessibility of spaces on behalf of, for example, all wheelchair users. PathVu[80] gathers and visualizes sidewalk information, but does not create a pedestrian network with the result or handle the complexities of pedestrian cost function definitions. Soundscape[79] models pedestrian experiences for the blind or otherwise vision-impaired, developing novel means of describing paths and the features along them, but is spatially limited in scope. World-scale mapping service like Google Maps™, Apple Maps™, and Bing Maps™ offer pedestrian and multi-modal routing options, but do not take into consideration primary pedestrian pathways or virtually any accessibility concerns.

The overriding theme of modeling and providing map information to pedestrians is inadequacy. No existing maps can demonstrate wide availability, adequate modeling of pedestrian mobility, a descriptive underlying pedestrian network, and meeting pedestrian informational needs (A to B routing, multi-modal trips) simultaneously.

1.4 THE FOCUS OF THIS DISSERTATION

The inadequacies of current pedestrian mobility models and user-facing applications demonstrate an unmet need regarding basic information required by the public: can I get from one public space to another in a reasonable amount of time and plan my trips accordingly? The inability to answer this question limits not just personal decisions on planning trips or outings, but also place limitations on investigating inequities in accessibility, infrastructure, and demographics served by public investments. Furthermore, the lack of rigor, data, and understanding of pedestrian mobility and relevant infrastructure raises fundamental questions about societal priorities, including inclusiveness in public spaces, ownership of data regarding public spaces, community planning and organization, and global warming, all of which impact and are impacted by increasing (or decreasing) pedestrian activity.

This dissertation lays out the needs and challenges in creating the infrastructure and user-facing tools necessary to understand the pedestrian environment in terms of individual pedestrians, which can be summarized as (1) a need to tackle all informational needs simultaneously to discover “unknown unknowns” as soon as possible, (2) understanding the distance between current data and approaches and what pedestrians actually need to know, (3) probing the diversity of pedestrian mobility concerns far more deeply, (3) defining data standards and exploring means of gathering data on pedestrian networks, (4) defining the problem of a pedestrian cost function for automatic routing and the form an adequate cost function will need to take, (5) opportunities to analyze the pedestrian network for public advocacy and planning purposes, and (6) the issue of data provenance regarding public pedestrian spaces. The experiences of the author in developing the AccessMap application for cities in western Washington will be used to demonstrate the extent of the technical barriers that have stymied adequate modeling of pedestrian needs, key insights into what future and more adequate pedestrian mobility models should look like, and a technical description of the approaches taken in building a city-scale, interactive, individualizable, accessible pedestrian network map.

The layout of this dissertation is as follows. Chapter 2 discusses the necessity of gathering and deploying public-facing pedestrian networks at scale in order to adequately understand the challenges in modeling pedestrian mobility. Chapter 2 also discusses and justifies the use of AccessMap as testbed for discovering and addressing these challenges. Chapter 3 discusses pedestrians’ current relationship with information about their environment and demonstrates informational inadequacies (gaps) using past work as well as original work surveying methods used by pedestrians to gather path information. Chapter 4 dives more deeply into how pedestrian preferences have been understood in prior work and presents original work on studying pedestrian preferences on environmental features using a Likert scale. Chapter 5 begins a discussion of the underlying data gap that undergirds the inadequacies in pedestrian-relevant information, beginning with the lack of a data standard that describes the fundamental features of how a pedestrian may traverse a public space. Chapter 5 also discusses the problem of data availability, as basic elements of pedestrian networks are usually not collected at all, and when they are collected are only available in non-standard and wildly different formats. Chapter 6 discusses the processes used to create an accurate and useful pedestrian network for AccessMap from Seattle agency data sources, demonstrating the technical work necessary to derive pedestrian networks from even data sources that appear ideal for a pedestrian network. Chapter 6 extensively discusses the extrapolation of network features from inherently non-networked data sources, pitfalls encountered, and strategies that may generalize to the more common formats in which pedestrian data are published. Chapter 7 explores an alternative means of acquiring pedestrian network data in a public way using crowdsourcing with OpenStreetMap, specifically as part of the OpenSidewalks project. The challenges, opportunities, and concrete mapping results of these endeavors are enumerated at length. Chapter 8 is dedicated to a unique challenge in creating an adequate pedestrian network: the estimation of incline (grade or steepness) at the scale of entire cities. Chapter 9 discusses a proposed standardization effort for pedestrian network data, the OpenSidewalks schema, and its relationship with OpenStreetMap, AccessMap, and other projects. Chapter 10 discusses the many, often unexplored, of describing pedestrian concerns with a cost function approach, as well as practical software engineering efforts to meet those needs. Specifically, two software packages created to support AccessMap are discussed: (1) Entwiner, a transportation network exchange format generator and API, and (2) Unweaver, a flexible, research-oriented routing engine. Chapter 11 discusses outcomes and the future potential for analyzing pedestrian networks for a variety of purposes, including successful lobbying efforts and understanding individualized accessibility at the scale of entire neighborhoods and cities with a new metric, SidewalkScore. Chapter 12 discusses philosophical and practical challenges of managing pedestrian data, which are inherently of public interest and often collected with public funding, and managing other stakeholder concerns. A data commons model is suggested and discussed. Chapter 13 covers a technical overview of the AccessMap project to assist with reproducibility and discuss the open source projects written during its creation.

CHAPTER 2. WE MUST IMPLEMENT AND APPLY PEDESTRIAN NETWORKS AT SCALE IN ORDER TO ENUMERATE AND EVALUATE KEY CHALLENGES

2.1 BACKGROUND

Due to the many challenges covered in the introduction, meeting the informational requirements of pedestrians will require simultaneously addressing a wide range of scientific and engineering challenges. Because this is the first known attempt to address all of these challenges at scale, we discovered several technical barriers to meeting these requirements. Said barriers may also be entirely contextual, such as attempting to conform pedestrian expectations with the structure of Dijkstra's Algorithm of shortest path finding in the context of limited data, so these challenges must be met simultaneously via prototypes and iterations within a joined context. In short, the problem domains should be explored in the context of their uses and connections, not in isolation. For example, knowing that manual wheelchair users prefer to avoid hills would measure a decontextualized preference, but does not realistically describe the trade-offs described in informal interviews. It may be that a portion of wheelchair users prefer to avoid hills near a particular incline, but only if it requires less than a five minute detour. Another portion of wheelchair users may prefer to avoid hills of a specific incline but only if it requires less than a five minute detour and there are no flat areas to rest along the hilly path every two minutes. Designing and implementing an informational system to model those preferences and an interface in which a pedestrian can efficiently communicate those preferences with a map will reveal unique challenges that will shape the discussion of what it means to adequately meet pedestrian informational needs. There are many such trade-offs to weigh, including many that are currently unknown, and only a live testbed will reliably tease apart what can be reported in a controlled setting versus what is truly useful and accurate in real-world conditions.

2.2 AccessMap as a testbed for the challenges of pedestrian networks

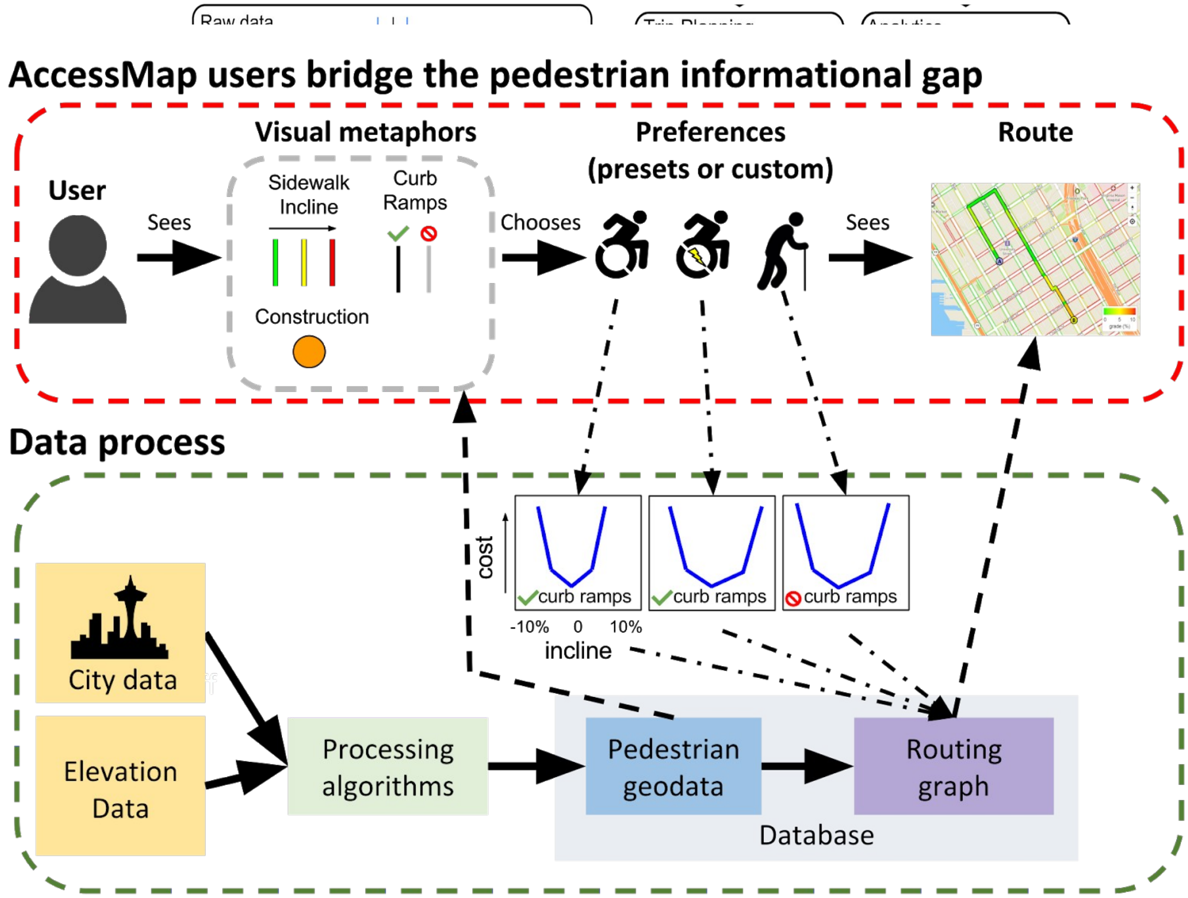


Figure 2.2: **How AccessMap addresses the pedestrian information gap via user experience and processing of open data sources.**

The workflow in the top panel illustrates a user's experience exploring AccessMap and via simple, consistent visual metaphors for pedestrian accessibility. Users can plan trips by: (1) selecting presets for common use cases, or (2) adjusting the routing settings manually to retrieve a trip plan customized to their needs. The workflow in the bottom panel illustrates how one-way data integration processing powers the AccessMap user experience. Raw data, first processed and imported into a single data format, are then converted to a routing graph in order to service custom routing requests.

dial in difficult-to-communicate settings like personal maximum uphill or downhill inclines. The routing engine is a separate microservice (Unweaver) we created that allows for arbitrary cost function definitions and web API generation in the Python programming language, allowing for rapid prototyping and study area-specific implementations when testing new data sources or approaches. The data displayed on AccessMap conforms to a public standard and is the same data used for routing, creating an opportunity to inspect and understand exactly why a route was chosen for an individual: clicking on any network segment, including a route, will bring up the infrastructure data used in the routing engine. As a public standard, the data defined for AccessMap can be reviewed for potential problems or limitations and extended or modified as needed. The data used by AccessMap comes from multiple sources and differs by city, with the data flows representing opportunities to understand the sufficiency and accuracy of common pedestrian data formats as published by agencies as well as crowdsourcing platforms. Finally, the routing engine and its data format are amenable to asking analytical questions of the pedestrian network, providing an opportunity to advocate for infrastructure improvements or otherwise better understand the role that infrastructure plays regarding pedestrian accessibility. Therefore, every element of AccessMap can be used to test or inform the grand challenges facing pedestrian information sources.

2.3 ACCESSMAP AND THE CHALLENGE OF SCALING

Personalized accessibility maps have been created in the past [35,37,41], but all have been limited by the scale of both attributes considered and the area served. Consequently, they served a spatially-limited audience and were often targeted at specific research questions without being a general-purpose tool for navigation. Some of these maps have ceased to function or are no longer available, highlighting another aspect of scaling: sustainability and robustness to funding limitations.

AccessMap operates at city and regional scale, providing the opportunity to address unique challenges: extracting reliable pedestrian network information from agency data, shared data standards, crowdsourcing pedestrian networks for large urban spaces, usage patterns and UI/UX flows, the scalability of flexible routing algorithms, visualizations, and urban analytics, and facilitating communication between stakeholders.

AccessMap currently operates based on a one-way flow of information, processing municipal or OpenStreetMap data into a routable graph and data that is visualized on its map. Due to this unidirectional information flow and lack of data and data standardization, expanding AccessMap to new regions will require either new and unique data processing challenges for each new agency data source or a concentrated mapping effort to crowdsource a pedestrian network. The awkwardness and labor required to process these data into a useful, standardized format are are discussed in Chapter 6.

While municipal data sources would benefit from reports of inaccuracies by motivated end users, this raises another challenge: no existing system can communicate geodata back to any city. To communicate the full extent of inaccuracies would overburden municipalities with technical overhead given their lack of data standards and tooling. For example, sidewalks are often inventoried as part of constrained legacy asset management systems; incorporating new geometries or changes would impose significant time and monetary overhead. Furthermore, municipalities open themselves to liability based on the data they publish and have no system in place to verify data submitted by users. Finally, multiple overlapping agencies all have an interest in accurate sidewalk data but must store private information, such as personally identifiable information on citizens, preventing submission of data back to municipalities and regional authorities. Thus, there is an unmet need for a system that coordinates these data across municipalities and other stakeholders.

To these ends, the AccessMap project undertook an effort to develop a sustainable data structure for pedestrian geodata and tooling to facilitate conversion of existing geodata. This work led founders of the AccessMap project to initiate the OpenSidewalks project, dedicated to advocating functional standards for pedestrian routes in OpenStreetMap (OSM), a crowdsourced mapping platform. OpenSidewalks developed and improved tooling to facilitate the import of municipal data into the more maintainable and flexible OpenStreetMap data format. The OpenSidewalks project therefore advocates for OpenStreetMap as a data commons for coordinating pedestrian data.

In combination, the AccessMap and OpenSidewalks projects cover nearly the entire gamut of attempting to provide fundamental information about pedestrian networks that are normally missing: data schemas, populating databases with compatible data from disparate sources, the self-collection of pedestrian network data, the formulation of path finding in terms of a shortest-path cost function, custom software to address all of these technical challenges, an investigable user interface, and large-scale analysis of pedestrian networks. Attempts to directly address these problems simultaneously revealed challenges and inadequacies that might otherwise be taken for granted, such as the data format of a sidewalk network, the appropriateness of Dijkstra-based graph traversal for pedestrian needs, the availability of appropriate shortest-path finding software, and even basic questions about what information pedestrians prefer to know in advance of a trip.

CHAPTER 3. PEDESTRIAN ACCESSIBILITY TOOLS PROVIDE INADEQUATE INFORMATION FOR PEDESTRIAN NEEDS

3.1 INTRODUCTION

Travel and access to transportation has changed significantly over the past two decades, giving travelers an unprecedented discoverability of available routes and services and offering travelers choice in selecting preferred travel options. Mobility applications (like OneBusAway[22] and NextBus[81]), and automated routing services (like Google Maps™, Apple Maps™, and Bing Maps™) are transforming not only travelers' access to multi-modal travel information, but how people actually use [65] and experience[2,54,61] travel overall. In turn, more people are accessing more destinations than ever before. With these advancements, applications serving information about travel services and environments play an increasingly important role in how people move around their communities, affecting access to employment, education, business opportunities, and medical services. However, one demographic yet to benefit in the same way from these mobility trends are travelers with accessibility requirements. Travelers with disabilities remain underserved due to pressing informational gaps and unavailability of data about specific attributes of the travel environments. Pedestrian environments are particularly challenging because unlike cars in roads, that uniformly need standard-sized lanes but few additional requirements, pedestrian needs and preferences in street environments are personal, and yet there are no consistent ways or standards for expressing what infrastructure pedestrians will find in these environments [7].

Since pedestrian environments link all transportation options to origin and destination points, this information gap is hard felt by travelers with disabilities pursuing independent travel. Accessible pedestrian navigation and routing remains vague, understudied, unevaluated, and unaddressed in mass produced mobility applications.

3.2 PEDESTRIANS FREQUENTLY RELY ON MANUAL INSPECTION OF DESTINATIONS PRIOR TO TRIPS

3.2.1 *Pedestrian information gathering habits user study approach*

Studying the informational requirements of pedestrians can and has been undertaken in many forms, including surveys[34] and accompanying pedestrians on trips[37]. However, a more fundamental analysis should begin with understanding what kind of information is currently sought out, the formats of consumption, and how frequently such resources are used. As part of an initial presurvey about pedestrian informational preferences (more in Chapter 4), study participants were asked to share the means by which they gather information prior to a pedestrian trip. Answers included getting help from a friend, using static (e.g. PDF or paper) maps, street-level imagery like Google Street View™, or web maps with automatic trip planning (like Google Maps™).

Mobility applications and other methods by which pedestrians can gather information prior to a trip may not serve pedestrians well: current trip planning tools may equip travelers with disabilities with incomplete, missing, or inappropriate data, making it hard or impossible to make independent, informed travel decisions. In addition, overestimating constraints may restrict viable and preferable options from consideration for a given mobility subpopulation. Aside from providing insights into trends and differences between pedestrians of different mobility types (individuals with different self-defined mobilities were queried), the option of street-level imagery is potentially of most interest, as such a tool is effectively the same as a preview of the actual space in question. If more automated options (maps) contained relevant and sufficient information, a street-level imagery tool like Google Street View™ could be assumed to be unnecessary. Therefore, significant or increased use of street-level imagery suggests an informational gap between pedestrians and web maps intended to accommodate for pedestrian needs.

3.2.2 *Pedestrian information gathering habits user study results*

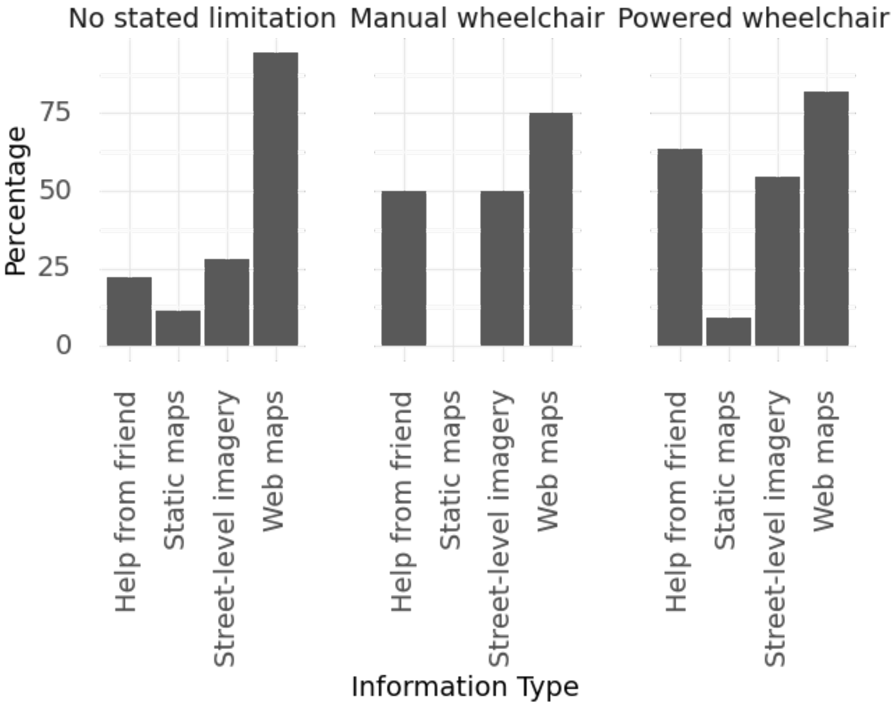


Figure 3.1: **All pedestrians make significant use of street-level imagery when planning a pedestrian trip.**

Wheelchair users making greater use of both help from friends and street-level imagery than pedestrians reporting no mobility limitation. Street-level imagery and friends may address informational gaps between what environmental features pedestrians find important and the information provided by web map routing services.

The results shown in Figure 3.1 demonstrate the frequent use of multiple tools to plan a pedestrian trip across populations, particularly the wheelchair user populations, including automatic trip planning with web maps and street-level imagery. This result is consistent with that suggestion that automatic trip planning applications lack relevant information for large subpopulations.

50 percent or more of wheelchair users indicated the regular use of street-level imagery or getting help from a friend prior to planning a pedestrian trip. These same disability populations also reported less frequent use of automatic trip planning than individuals that state no mobility limitation. When participants were asked if they would feel confident to take the trip based only on Google Maps'™ directions instructions without investigating street-level imagery, one

participant related that they would rather send out a scout in advance to assess the route for them than attempt the travel on their own with no additional information. These results suggest that pedestrians with disabilities are cognizant of pedestrian routing applications and their benefits, but also of the information they are unable to provide, and use combined and often laborious processes to ensure a safe, accessible trip.

3.2.3 *Variability in disability population preferences may be addressable by individually-customizable accessibility maps*

Following up on the presurvey questions, we designed and implemented semi-formal interviews and a mapping activity with pre-designated start and end points for a pedestrian trip. When asked to plan a route in an unfamiliar neighborhood with any mapping or routing technology of their choice, all participants chose to use 'Google Directions' to plan their route and then immediately went to an additional tool (typically "Google Street View", or other imagery tools) to visually inspect the built environment at that location and ascertain visually whether it suited their needs. This latter visual inspection process took anywhere from 90 seconds to 9 minutes in the laboratory setting, and users reported sometimes spending up to 4 hours planning a single trip with a host of other tools, like Google Earth, to assist with visual inspection.

In the next segment, participants were asked to plan and evaluate routes using three investigatory methods: (1) Google Maps™ Directions, (2) manual investigation of AccessMap, an accessibility map that includes sidewalk, steepness, and curb ramp information, and (3) an automatically-planned trip using AccessMap using personalized settings. Following the generation of each route, participants were asked to rate their confidence in the appropriateness and usefulness of each route on a scale from 1 to 5, with 1 being low confidence and 5 being very high confidence (Figure 3.2). Manual trip planning with AccessMap entailed a maneuver-by-maneuver exercise wherein a given route was recorded by researchers as participants simulated traveling along potential paths, finally resulting in a single canonical path from start to end point. Many participants described bet-hedging approaches, stating that they would attempt one path tentatively but use a backup if it was untenable. For automatic trip planning with AccessMap, participants were asked to enter their personal navigational preferences regarding uphill steepness (2-15 percent), downhill steepness (2-15 percent), and curb ramps (required or not) and instigate a trip with the designated start and end points.

Figure 3.2: **Five-point ratings of the appropriateness and usefulness of five different styles of pedestrian-serving maps based on in-person follow-up interviews.**

Points represent mean average score and lines represent a standard deviation on each side. Points with no standard deviation line were unanimous. Scores for the AccessMap test case, in which personal preferences on steepness and the availability of curb ramps dictate route selection, repeatedly scored high for both usefulness and appropriateness for all mobility categories studied.

Perhaps unsurprisingly, wheelchair users were consistently more confident in both the usefulness and appropriateness of a trip planned using a map (AccessMap) that includes more information about the pedestrian environment (inclines, curb ramps, sidewalks) (Figure 3.2, middle and right panels). Wheelchair users reported greater confidence when presented with more information (AccessMap vs Google Maps™). There were 18 participants reporting no mobility limitation, 4 reporting the use of a manual wheelchair, and 11 reporting the use of a powered wheelchair.

Individually-customized route planning provided an opportunity to accommodate intra-group variation, and wheelchair users were consistently slightly more confident in a route that was planned automatically for them after inputting their personal settings than in their manually-planned route. Trip-planning services that account for the variability within and between pedestrian populations with and without disability designations may therefore obviate the current need for street-level imagery audits prior to confidently starting a pedestrian trip.

CHAPTER 4. RESEARCH ON THE RELATIONSHIP BETWEEN PEDESTRIANS AND THEIR ENVIRONMENTS IS UNDER-RESARCHED AND FREQUENTLY STEREOTYPES

The challenges of meeting the informational requirements of people with disabilities in diverse travel environments have been the subject of multidisciplinary efforts and research trying to understand and support navigation using assistive technologies. The main challenge in this body of work is how to bridge a fundamental informational gap: what information matters to individuals with accessibility concerns to improve independent navigation experiences?

In the complex range of systems and support services currently in place to address the needs of people with disabilities, taking individual variability into account is driving individualized plans in many realms (education, assistive technology, long term care, etc.). Yet, in access to travel, assistive technology approaches have fundamentally neglected to acknowledge this variation and rather find motivation to coalesce around grouped mobility profiles, often based on assistive devices. While not explicitly stated, this is potentially motivated by the assumption that building assistive navigation devices or technologies could be facilitated through building dedicated group-specific tools. With the level of customization and personalization that current technologies allow, we question the extent to which assumed technological hurdles should drive assistive technology design decisions.

We explore current practices around the logic of difference in navigational preference research. Studies to assess navigational preference frequently invoke assumed notions of difference among disability groups without probing for granular differences within and often between groups. Though unintended, this approach may function to artificially enforce homogeneity along some dimensions of preference and heterogeneity along others in the study populations. In this work we argue that the navigational preferences of people with and without disabilities are personal and potentially situational properties – and that people’s mobility preferences may be better aligned across other divides than type of disability or use of a specific mobility device. Questioning the assumed groupings used to segment study populations may help us better understand the contingencies that shape routing and navigation decisions for all pedestrians.

4.1 PEDESTRIAN PREFERENCES ARE PERSONAL, BUT PEDESTRIANS ARE STEREOTYPED BY DEVICE

Typically, researchers have interrogated questions of informational requirements for accessibility in the context of specific disability groups in order to model the needs of people with different disabilities in different travel environments[25]. For example, prior work had focused on people with mobility limitations in outdoor environments[18,34,41,42,47], visual impairment in indoor environments[1,11,52,53,69], visual impairment in outdoor environments[4,8,10,27,33,52,57,66–69], hearing[4] or cognitive disabilities[4,12], and older adults or people with multiple disabilities[3,56,62]. One recent approach advocated for synthesizing all preferences together towards a universally acceptable navigation technology[29]. Assistive devices have often served as a means of subdividing or uniquely identifying disability groups in this research and many studies have relied on low sample sizes due to the nascent nature of this research and the many application domains of accessible technology.

In 2016, the Accessible Transportation Technologies Research Initiative or ATTRI (a joint USDOT initiative, co-led by the Federal Highway Administration (FHWA)), produced an extensive stakeholder report. Unlike prior work with relatively small sample sizes, this report surveyed travelers with disabilities' needs and barriers among 1203 individuals, identifying and ranking four categories of interest to the studied population (“Information”, “Options”, “Assistance”, and “Access”[51]. The results showed that participants ranked highest the needs and barriers in the ”Information” category above any other need or barrier category (followed by travel “Options”, “Assistance”, and “Access”). The “Information” category appeared in 74 percent of all cited needs and 48 percent of all cited barriers by the entire surveyed population. Specifically, participants overwhelmingly prioritized “Information” needs and barriers over both “Access to Technology” and “Physical Access” to environments[51]. Participants also showed they are savvy about the natural next development in technology: they want technology to bring all of the information and travel options together on a common platform. Participants discussed the need to plan trips across a range of travel modes and environments. Significantly, they stated the need for personalization based on their preferences regarding accessibility requirements, comfort, cost, and convenience, but identified the discomfort in allowing a mobility application or a transportation agency to be aware of their specific mobility requirements. They explicitly considered this type of information as personally identifiable information (PII), implying that people with disabilities consider their preferences to be unique to them, rather than self-identifying as aligned with a particular disability or assistive-device group.

This result raises important questions regarding the validity of the ways accessibility is assessed for mobility-constrained populations where accessibility is often treated as a monolith[45] or grouped by population attributes such as by disability or assistive device[34,47,77]. For example, an application focused on wheelchair users may not distinguish between manual or powered wheelchair users, who exhibit distinct informational preferences[77].

4.2 PEDESTRIAN PREFERENCES ARE DIVERSE

Because pedestrians consider their informational requirements to be of potential to personally identify them, there may be significant diversity not just between pedestrians in assistive device categories, but within them and along unexplored axes of variation such as comfort level, athleticism, or other unmeasured variables. Realizing that variation exists within and between typical disability groupings and that accessibility requirements are highly personal information, some assistive technologies have promoted personal customization over rigid groupings[7,18,41]. Perhaps, these approaches suggest, the informational requirements of manual wheelchair users with extensive experience and athleticism differ from individuals who have only recently started using manual wheelchairs: existing data rarely distinguishes these populations nor captures the degree to which users of a particular device may disagree with one another, but both would be summarized as “wheelchair users” by device. Meeting the informational requirements of people with mobility limitations will require questioning and probing current stratification, investigating prioritized attributes and attribute granularization: on which environmental attributes is there intra-group disagreement and are there more specific questions that should be asked? Given that there is latent diversity in mobility preferences to the extent that it is considered PII, it can be hypothesized that users of a given assistive device are not sampled from a homogeneous population, but multiple distinct populations that may be isolated and prioritized for assistive technological interventions and needs assessments. The identification and extraction of such subpopulations will require datasets with high granularity of navigation preferences and approaches to identify meaningful clusters.

4.2.1 *Pedestrian diversity user study approach*

A primary objective of this study was to explore intra-group and inter-group variability in informational requirements for three pedestrian subpopulations: individuals with no stated mobility limitations, manual wheelchair users, and powered wheelchair users. To do so, we carried out a study consisting of an online survey that collected demographic information, personal mobility information, and built environmental informational preferences regarding pedestrian trip planning. Respondents were recruited through mailing list connections, local accessibility advocacy groups, and university outreach, followed by snowball sampling. 15 respondents self-identified as having a mobility limitation and 18 self-identified as having none. Of those self-identifying as having a mobility limitation, 4 used a manual wheelchair as their primary device and 11 used a powered wheelchair as their primary device. Participation was not limited by assistive devices used and some participants were recruited who used prostheses, oxygen tanks, or canes, but were excluded from this publication due to low sample size (one to two for each device).

Participants were asked to rate the importance of foreknowledge of seven environmental attributes relevant to pedestrian mobility on a five-point Likert scale ("Not Important", "Slightly Important", "Moderately Important", "Important", "Very Important") before starting a pedestrian trip. The environmental attributes queried were (1) construction that may block a path, (2) cross-slope (steepness orthogonal to the path of travel), (3) the presence of curb ramps, (4) slippery surfaces, (5) steep downhill inclines, (6) steep uphill inclines, and (7) unpaved surfaces.

4.2.2 *Likert scale self-assessments reveal widespread, granular variation in informational requirements*

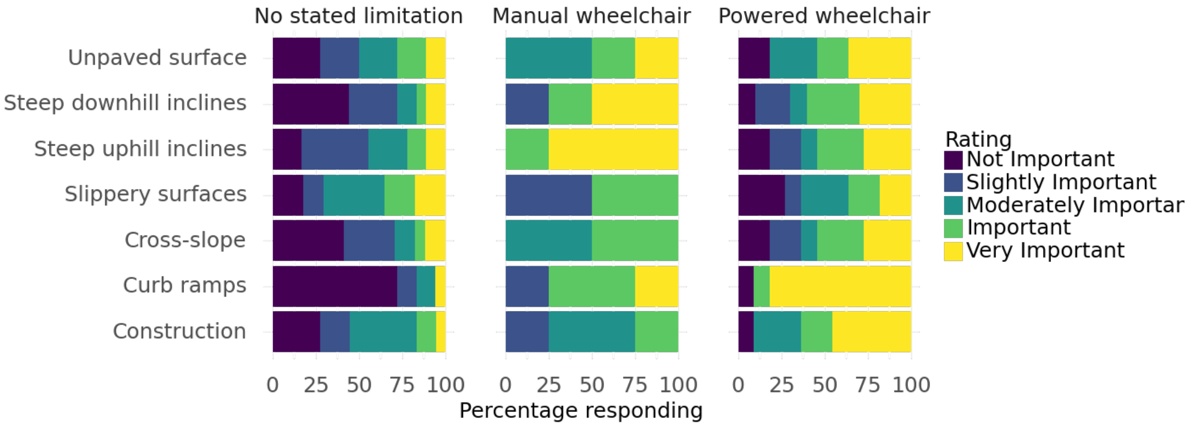


Figure 4.1: **Five-point Likert scale ratings of the importance of foreknowledge of pedestrian environmental features by pedestrians.**

Pedestrians were segmented by self-ratings of mobility limitations and by devices used: pedestrians reporting no mobility limitation, manual wheelchair users, and powered wheelchair users. Each population possessed distinct preferences regarding nearly every environmental feature and many environmental features elicited significant intra-population variation in ratings such as powered wheelchair users' ratings of cross-slopes.

Study participants exhibited wide variations in how they rated the importance of information about the built environment prior to a pedestrian trip (Figure 4.1). Study participants with no stated mobility limitation acted as a baseline comparison for a normative pedestrian population, and mapping services target this population. Nevertheless, such participants exhibited non-uniformity in their ratings of the importance of environmental feature foreknowledge. For example, for all but one (curb ramps) of the environmental features, participants with no stated mobility limitation there was no single Likert scale rating receiving more than 50 percent of responses. For example, no response for preferences on unpaved surfaces exceeds 30 percent or is below 10 percent, suggesting a wide and roughly flat distribution. In contrast, foreknowledge of curb ramps was heavily biased towards low-importance ratings. For individuals with no stated mobility limitation, all other environmental features fell between these two extremes: somewhat biased distributions. Together, these results suggest that the pedestrian population typically treated as normative, with no self-identified mobility limitations, has informational requirements that are not met by mainstream map services.

The environmental feature ratings provided by manual wheelchair users (Figure 4.1, middle panel) similarly exhibit somewhat wide distributions of responses. Due to the number of study participants using manual wheelchairs, strict conclusions about the shape of the distributions are difficult to identify with certainty, but cases of agreement and disagreement within can still be identified. For example, all respondents indicated either a rating of "Important" or "Very Important" regarding steep uphill inclines, a result consistent with previous summary statistics regarding this population. However, prioritization of information about curb ramps, slippery surfaces, steep downhill inclines, and construction each resulted in at least one "Slightly Important" rating, from a variety of individuals, while others indicate "Important" or "Very Important". These results suggest that a "one-size-fits-all" approach, even within a commonly-identified disability population, will fail to capture much of the population. For example, some manual wheelchair users state a strong preference ("Important", "Very Important") for knowing the availability of curb ramps, while others give it only a "Slightly Important" rating. This reflects the authors' experience in informally interviewing manual wheelchair users: not all manual wheelchair users state a strong interest in knowing when curb ramps are available, relying on jumping curbs or locating a driveway when curb ramps are absent. Assistive technologies that assume all manual wheelchair users require curb ramps could therefore fail to represent the preferences of many manual wheelchair users, and would moreover, refrain from offering route options that are appropriate for them (at the potential risk of finding no routes at all).

The distribution of responses from powered wheelchair users' (Figure 4.1, right subplot) show several wide and flat distributions and potentially bimodal distributions (disagreement) within the population surveyed, with only one case of near-consensus: foreknowledge of curb ramp availability. For each feature, at least 9 percent of respondents indicated that the feature is "Not Important" to know prior to a pedestrian trip and at least 18 percent indicated that the feature was "Very Important". For cross-slope, slippery surfaces, downhill and uphill inclines, and unpaved surfaces, respondents show wide ratings of importance with no clear bias in the distribution of responses. Prioritizations regarding cross-slope could potentially even exhibit bimodality, though a higher sample size would be required to reach a firm conclusion. Consequently, a "one-size-fits-all" approach to uphill incline in an assistive technology may also miss a significant portion of powered wheelchair users. An accessibility map with a single powered wheelchair setting, or even one that lumps together powered wheelchair and manual wheelchair users, will likely rule out paths as inaccessible that might be preferable for many powered wheelchair users.

Independently querying uphill or downhill steepness (grade) informational priorities was a deliberate design choice made to probe whether there are disagreements or independent trends in whether one is traversing up a steep path or down a steep path, which likely engage different risks and challenges among wheelchair users. In informal interviews, manual wheelchair users have indicated that uphill travel is strenuous and may represent a hard mechanical barrier, while steep downhill inclines may represent a safety hazard, with a fear of losing control of the device. Consistent with other work[34], our study shows that manual wheelchair users show a greater concern for information about inclines than do powered wheelchair users.

One potential hypothesis for the variability seen within pedestrian populations is limitations in self-assessment: i.e., that manual wheelchair users agree more than our data suggest in reality, but the self-assessments introduce noise in responses. However, three manual wheelchair users described their deprioritization of curb ramp information by relating their approaches and experiences: two indicated that they would attempt jumping most curbs to save time and one that they expect driveways to be available as an alternative. One manual wheelchair user described traversing raised curbs by exiting their wheelchair, dragging it over the curb, and re-entering their wheelchair once on the sidewalk. These anecdotes also suggest an important future topic of study: the potential trade-offs between pedestrian trip considerations, such as the time of travel vs. partial barriers.

Therefore, it is likely that much of the extent of fine-grained agreement and disagreement within and between these pedestrian subpopulations demonstrate real variation and that there are likely undiscovered and meaningful alternative segmentations of these populations. If identified, alternative segmentations may provide a more appropriate means by which to serve pedestrians of all abilities, whether it is through more detailed accessibility analyses of pedestrian spaces or the development of assistive technologies like entirely customizable digital maps.

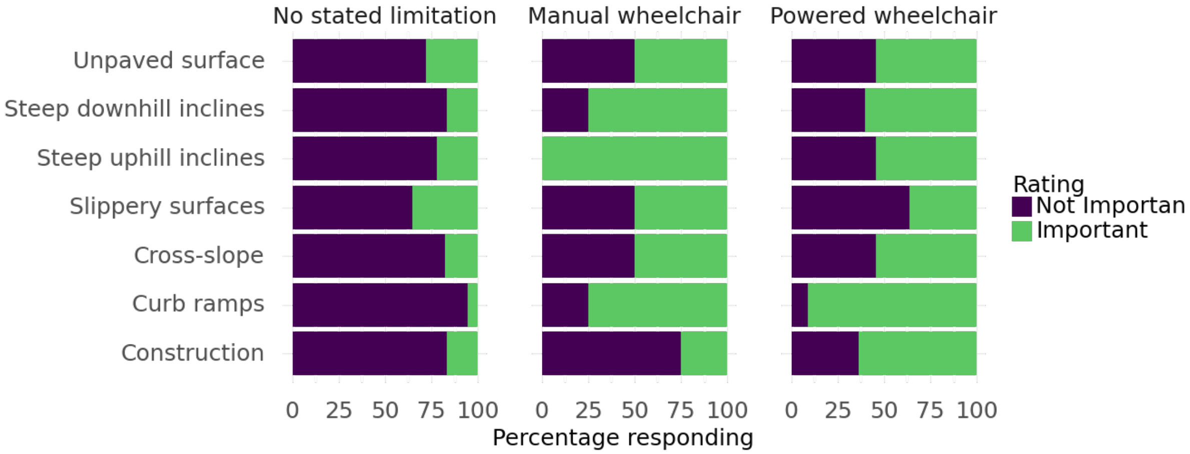


Figure 4.2: **Environmental feature foreknowledge ratings converted from a Likert scale to binary scale.**

A binary scale may hide meaningful variation that could inform evaluations of accessibility and the design of accessible technology.

Previous work in studying the preferences of mobility-limited pedestrians have often requested binary responses to features of the built environment such as, "is this feature significant?". In comparison to a more granular approach, such as the Likert scale used in this study, a binary scale will overlook variation in intensity of preference and potentially distinct subpopulations. In order to probe for potential impacts of losing granularity, we transformed our survey responses onto a binary scale: the "Important" and "Very Important" responses were converted into an "Important" class and all other responses were converted into a "Not important" class (Figure 4.2).

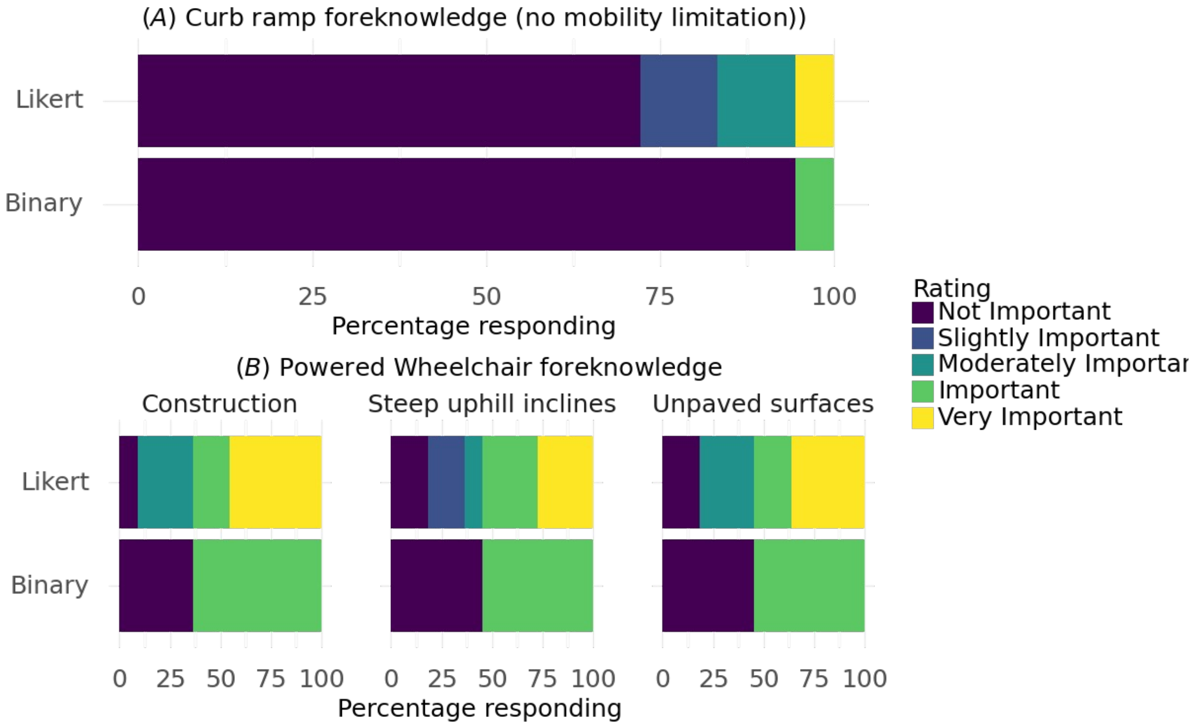


Figure 4.3: **Binary scale needs assessments overlook significant variation.**

(A) A Likert scale reveals slight preferences for foreknowledge of curb ramp availability in pedestrians with no stated mobility impairment. (B) Ratings of the importance of foreknowledge of three environmental features among powered wheelchair users appear similar on a binary scale, but a Likert scale reveals greater variation in relative importance that could otherwise be overlooked.

The (less granular) binary scale necessarily hides variation that could impact the design of accessibility maps, navigational aides or research into the accessibility of urban spaces by giving a misleading impression of certainty. For example, Figure 4.3A shows that about 28 percent of participants with no stated mobility impairment rated curb ramps as either “Slightly Important”, “Moderately Important”, or “Very Important”, but the binary scale would only identify 5 percent of this population as prioritizing curb ramp foreknowledge in any way. These results suggest that a significant population of pedestrians with no stated mobility limitation have at least some interest in curb ramp information, an intra-group subpopulation that could be missed by the (common) binary scale approach, but a pedestrian-centric map technology should potentially include this information even for pedestrians that state no mobility limitation.

In addition to masking granular priorities, a binary scale may hide the shape of variation or disagreement and the separability of potential subpopulations. For example, for powered wheelchair users, foreknowledge of construction, unpaved surfaces, and steep uphill inclines have similar (or identical) ratings on the simulated binary scale, but have contrasting distributions on a more granular Likert scale (Figure 4.3B). As expected, the binary scale hid the severity of agreement with the rating. For example, foreknowledge of unpaved surfaces have a larger proportion of the “Moderately Important” rating compared to steep uphill inclines, even though both are “Not Important” on a binary scale. Construction information presents a similar situation, where the binary scale rating hides a significantly-sized “Moderately Important” subset of powered wheelchair users.

Overlooking this granularity in disability population preferences has the potential to impact the design of accessible technologies or the extraction of meaningful pedestrian subpopulations. For example, the developer of an accessibility map could conclude from these features are similarly important on a binary scale, and their inclusion would serve similarly-sized populations, but if trends hold from our study, construction information would ideally be included before the other two features or be prioritized in the user interface.

In addition, the intra-group variation demonstrated in Figure 4.1 suggests an opportunity to further segment pedestrian populations using granular information that would be difficult when using a binary scale. On a binary scale, the decision to split powered wheelchair users along, for example, the dimension of steep uphill inclines may be valid (Figure 4.3B), as this population may be bimodal in its preferences, but not along construction preferences, which is more smoothly biased towards high ratings.

4.2.3 *Discussion*

The work presented here suggests a need for greater emphasis on intra-group disability population variation in both accessibility research and in the application of accessible technologies. A primary but insufficiently tested hypothesis raised in this work is that there exist distinct subpopulations within traditional disability categories that could be extracted with sufficient data. The fact that disability populations consider their navigational preferences to be PII is supportive of this hypothesis. Future research can build on these results to explore this topic more deeply and with greater specificity.

4.3 DIVERSITY IN MOBILITY IMPLIES A NEED FOR NEUTRAL PEDESTRIAN NETWORK DATA

Some maps or informational tools attempt to address diversity in pedestrian mobility by offering alternative “modes”. For example, Google Maps™ offers limited a “wheelchair accessible routes” mode for a subset of transit stations[72] and OpenRouteService offers a dedicated wheelchair pedestrian mode in European countries[77]. In addition, attempts to capture pedestrian diversity by offering a single alternative are common outside of the context of interactive maps, such as the identification or designation of dedicated ADA-compliant paths meant to cover a baseline or minimal level of accessibility.

Such “one-size-fits-all” approaches to “alternative” pedestrian mobility modes describe a path taken for a hypothesized most-constrained scenario, attempting to ensure that any pedestrian meeting a particular description (e.g., wheelchair user) will be able to make use of the path provided. While they may meet this goal, they do not capture the actual path that many pedestrian may desire. For example, an athletic manual wheelchair user may strongly prefer to climb relatively steep hills to save time, but would only be provided with a slower, longer path intended to meet a most-constrained mobility model. Clearly, a single alternative mobility mode is inadequate to capture variation and may even be less useful than a bicycle or unconstrained pedestrian mode for many pedestrians with mobility limitations.

To make matters worse, attempting to define a single alternative mode, such as “wheelchair mode”, removes the decision making power from the pedestrian attempting to traverse a space and places it in the hands of a third party that may not be effective or consistent at defining wheelchair accessibility. The decision to stereotype modes of pedestrian mobility therefore has data modeling implications: is it appropriate for the underlying data to mark a path as “wheelchair-accessible”? For example, OpenStreetMap defines many “tags” (key-value pairs communicating map data) under the “wheelchair=*” key, where a point or line may be defined as “wheelchair=yes” or “wheelchair=no”. Individuals mapping spaces in OpenStreetMap may therefore define a path as wheelchair-accessible or not from the other side of the world, attempting to decide on behalf of all local wheelchair users whether a given piece of infrastructure is usable, accessible, or preferable. Whereas “wheelchair=no” is suggested for raised curbs, many wheelchair users indicated in informal interviews that they would be happy to jump a curb to save time on a trip. Rather than embed an individual’s opinion in this accessibility data, a description of a curb and its height would be more flexible and could be used to model pedestrian diversity. The need for individually-conditioned interpretation of map data requires the collection of ostensibly neutral, on-the-ground information that can be interpreted as needed for pedestrians.

CHAPTER 5. THE DATA GAP: WHAT INFORMATION IS REQUIRED VS WHAT IS CURRENTLY AVAILABLE

5.1 NECESSARY PROPERTIES OF AN INCLUSIVE PEDESTRIAN NETWORK

Pedestrians' needs for a wide range of individualized information in order to reliably plan a trip and the current dearth of such information necessitates the creation of a detailed pedestrian network. But what should the schema of such a network entail? Pedestrian infrastructure like sidewalks have been published as points, lines, polygons, street metadata, and images in a variety of open datasets, but we must settle on a network-compatible format in order to rapidly interpret traversability. While sidewalk centerlines are a promising choice for data format, analogous to centerlines in street networks, sidewalks alone are insufficient to describe a pedestrian network and leave many ambiguous situations that require resolution. When the pedestrian network crosses a street, where does the sidewalk stop and the crossing begin? Does the path along a curb ramp constitute a sidewalk, a separate curb ramp feature, or an element of the crossing? Some potential answer to these questions will be elucidated later in this document, but we should first discuss what information that must necessarily be represented in order to build a fundamental pedestrian network for an urban space: sidewalks, crossing locations, curb ramps, and streets. In combination, these constitute the basic network that is sufficient for basic connectivity information. In addition, to be immediately useful for a wider range of pedestrians, this network should be enriched with metadata: inclines, marked / protected crossings, construction information, etc. An example schema is shown in Figure 5.1.

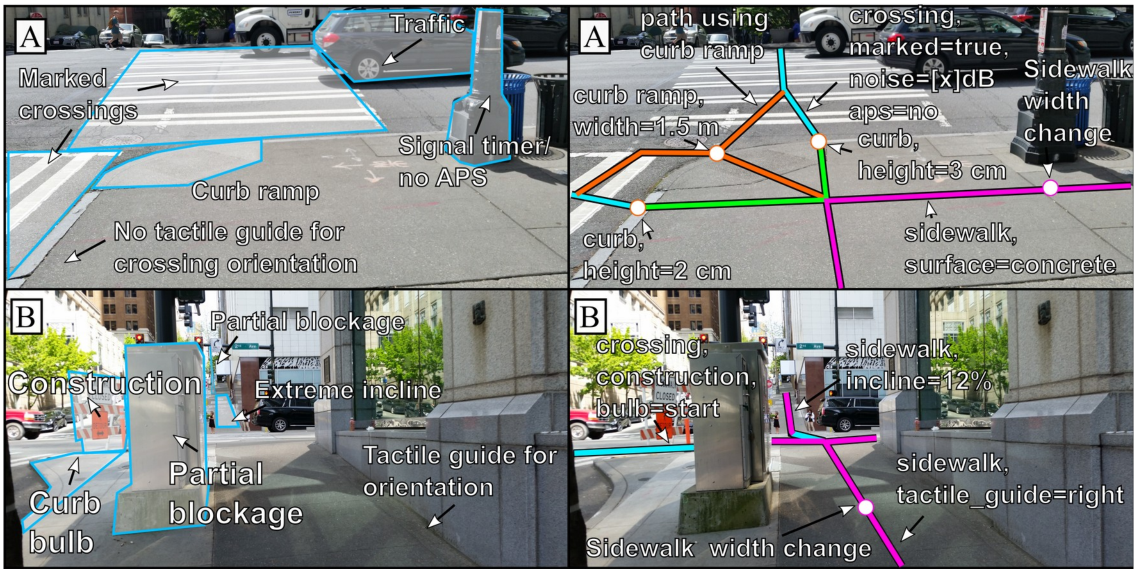


Figure 5.1: **Example schema of a minimal pedestrian network**

(A) (left) Features of interest at a crosswalk. (B) (left) Features of interest along an urban sidewalk. (A) (right) A street intersection, where street and pedestrian networks meet. A pedestrian seeking to cross the street may take several paths off of the sidewalk depending on their preferences and mobility constraints, and these should be represented (curb ramp paths and non-curb ramp paths). In addition, width information, curb height, the availability of an APS, and whether a crossing is marked are all important information. (B) (right) A primary pedestrian network example demonstrating a sidewalk width change due to an installation in the middle of the sidewalk. Multiple schemas could represent this situation, but all should include the effective path width.

5.2 A SURVEY OF POTENTIAL PEDESTRIAN NETWORK DATA SOURCES

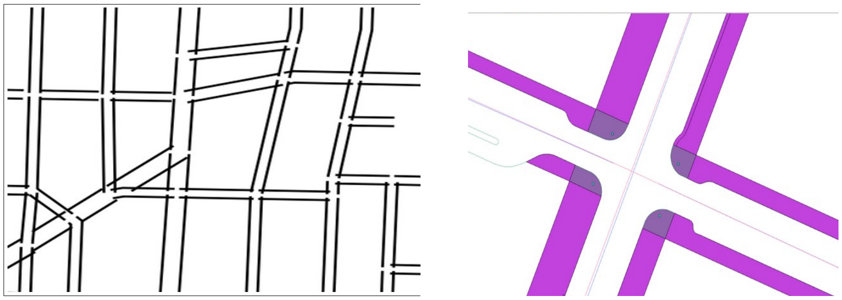


Figure 5.2: **Representative data types for agency sidewalk data**

Sidewalk data is typically available in linear (left) or polygonal (right) format.

While it is clear what information is essential in a pedestrian network and what some of the properties must be (e.g., an embedded graph representation), this kind of information is rarely published, and when it is published, it is in non-standard formats(Figure 5.2). Table 5.1 displays the results of a survey of major US cities as well as cities for which it is straightforward to identify pedestrian network data sources. Columns include the most basic information for describing an urban pedestrian network: sidewalks, crossings (marked and unmarked), and curb ramps. Crossings are distinguished by marked or unmarked because cities sometimes keep an inventory of marked crossings but not unmarked. Each row shows either P (point), L (line), A (area), or is left blank (no data) to indicate the format of data that is present for a given city and pedestrian network feature type. While the data must be transformed into a linear representation in order to be routable, a wide range of formats can be seen for each feature type and crossings are nearly always missing.

Table 5.1: Pedestrian network data published by US cities

Region	Sidewalks	Marked Crossings	Unmarked Crossings	Curb ramps
Seattle, WA	L	P		P
Denver, CO	L, A			P
Arlington, VA	L	L	L	A
Portland, OR	A			P
Chicago, IL	L			
Hartford, CT	A			A
Washington, DC	A			A
New York, NY	A			
Los Angeles, CA	A	P		A

CHAPTER 6. DERIVING A PEDESTRIAN NETWORK FROM AGENCY DATA: A CASE STUDY IN SEATTLE

6.1 BACKGROUND

The AccessMap project developed three strategies to increase the accuracy of parallel-offset sidewalk data: a heuristic decision tree, machine learning from crowdsourced data, and an algorithm that uses inferred street metadata to redraw more accurate offset lines based on geometrically buffering streets. The last method proved most useful for Seattle. The redrawing method requires only an estimate of offset distance and a flag for whether a sidewalk is present on the left or right side of the street (Figure 2.1). It could therefore be used to leverage datasets that store sidewalk information in street metadata. These efforts enable the translation of geodata in a variety of formats to a single format that is consistent, routable, and represents an intermediate solution in the path to a longer-term, shared data layer for pedestrian geodata.

Pedestrian ways that connect sidewalks are necessary for routing, but the locations of legal crossings in the Seattle metropolitan region, as in many other regions, are not explicitly cataloged. In Seattle, any intersection is a legal pedestrian crossing unless otherwise marked, but laws regarding crossings vary regionally and internationally, including allowing crossing at nearly any point on a residential street in many parts of Europe. As such, explicit, long-term storage of centerlines for all crossing locations may not be suitable for all locales. For the purposes of AccessMap Seattle, we algorithmically generate street-crossing centerlines that connect around intersections and then annotate them based on the existence of curb ramps at both ends. The result is a well-connected sidewalk-and-crossing pedestrian graph covering the entire Seattle area.

To enable accessibility-focused routing for pedestrians, we needed to assign attributes to each sidewalk and street crossing that can inform our routing cost function. Curb ramp information was derived from latitude-longitude coordinates published by SDOT and used to flag whether street crossings had curb ramps at both ends. However, these data suffer from spatial inaccuracy: curb ramps are stored as metadata by SDOT rather than as explicit latitude-longitude coordinates; as a result, the locations reported at data.seattle.gov are extrapolated from SDOT sidewalk locations, whose endpoints are incorrect. Consequently, AccessMap assigns curb ramp metadata to crossings through table joins and inferred metadata from SDOT in Seattle and uses spatial proximity in all other cases. Adding estimated inclines to paths is a substantial challenge due to the resolution and availability of elevation data and is covered in Chapter 8.

6.2 DERIVING CONNECTED SIDEWALK LINES

6.2.1 *Fixing sidewalk data in-place*

Figure 6.1: **A divide-and-conquer strategy for networking Seattle's sidewalk data in-place**

Seattle's sidewalk data has inaccurate endpoints, leading to centerlines with disconnection and intersection errors. A divide-and-conquer strategy was developed using PostGIS to separate the problem into blockfaces and then connect nearest neighbors.

To make our sidewalk denoising strategy generalizable, we categorized the noise in the Seattle datasets as city-specific noise to be cleaned prior to entering the data processing tool's workflow, and city-agnostic noise that our tool would address. The primary Seattle-specific source of noise identified was a systematic error at 'T' intersections, where three streets meet and one of the angles between them is large. At these intersections, inaccurate and large gaps were systematically generated during sidewalk extrapolation from by the asset management system used by the Seattle Department of Transportation. We connected the sidewalks intersections by identifying 3-way intersections with angles between 170 and 190 degrees and connecting the appropriate sidewalks within 100 feet. We connected 88.2% of the 4,352 'T' intersection sidewalks in the Seattle sidewalk data prior to entry into the data cleaning framework.

The initial step in our data processing framework is to connect sidewalk segments that should be connected in reality, but are disconnected in the data due to noise in the source data. The error in sidewalk locations can be large enough that endpoints erroneously appear on the wrong side of the street, appearing closer to neighboring sidewalk endpoints than to correct endpoints. As a result, proximity alone is insufficient to make a decision on whether two sidewalk segments should be connected. To address this, we developed a series of topological rules that first classify sidewalk endpoints by street blocks using street intersections before connecting sidewalk ends based on proximity (Figure 6.1). This approach relies only on prior knowledge of sidewalk location and street locations and can therefore be generalized to any city. Once connected, these sidewalks form subgraphs, typically representing the sidewalks connected around a block, but disconnected from one another (no across-street connections). The coverage of our algorithms was evaluated by the number of modified sidewalks relative to the total number in the data set. Our cleaning and preliminary data analysis showed that we edited 86.9% of the sidewalk corners at intersections and made 90.6% of sidewalk subgraphs fully connected within their block.

6.2.2 *A machine learning approach with crowdsourced labels*

While developing these heuristics, we also explored a crowdsourced (“citizen scientist”), machine-learning based method to tackle the problem of generating a connected graph for sidewalks. While we no longer use this method for AccessMap Seattle, it represents a potential path forward to resolve in a generalizable manner many other pedestrian data questions.

Through our efforts to improve sidewalk data, it became apparent that individuals could readily predict sidewalk connectivity from visual cues in web maps. This motivated two hypotheses: (1) that untrained users could provide a reliable ground truth dataset for comparing heuristics, and (2) given this labeled data and a set of features, we could apply supervised learning to predict sidewalk connectivity on a city-wide scale. A key advantage of this approach is scalability: errors in municipal data systematically vary from city to city, foiling a static heuristic; however, by learning the weights of universal sidewalk attributes from crowdsourced data, we could learn to fix sidewalk data on a regional basis from relatively few examples.

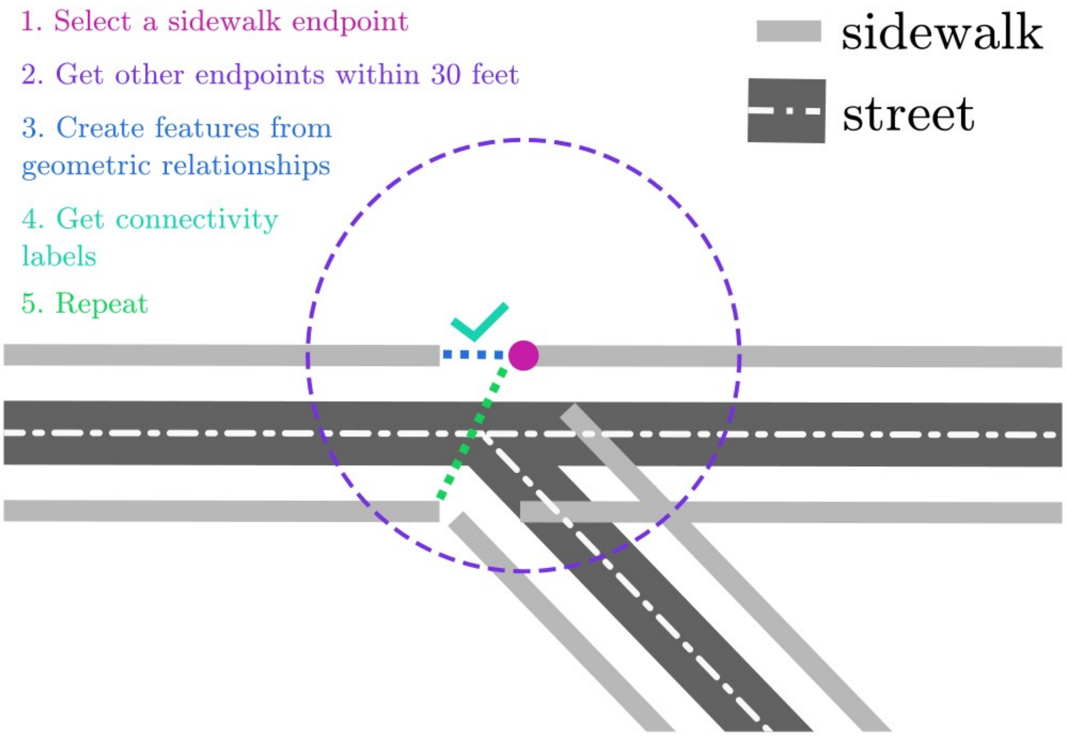


Figure 6.2: **Feature generation for the supervised learning of sidewalk connectivity in Seattle**

Seattle's sidewalk data has inaccurate endpoints, leading to centerlines with disconnection and intersection errors. In order to improve on a heuristic approach that used guessed-at spatial parameterizations and logic, we created a dataset of sidewalk end-end pairs along with features describing their relationship, such as distance, angle of approach, and whether a connection would require crossing a street line.

We posed the sidewalk classification problem for supervised machine learning thusly: given any two sidewalk endpoints of different sidewalks, should the two endpoints be connected (Figure 6.2)? The goal was therefore to predict a label (1 - connected or 0 - disconnected) for new examples based on a relatively small training set. Each learning example constitutes a relationship between two sidewalk endpoints and used spatial relationships as learning features, such as the distance between sidewalks, whether the sidewalks intersect, the angle of approach between sidewalk ends, whether sidewalks fall within the same city block, and whether a line connecting the sidewalk ends would cross a street. Non-spatial features included both sidewalks having the same sidewalk width or surface material. To reduce the problem size, we limited sidewalk pairs to those within 10 meters of one another.

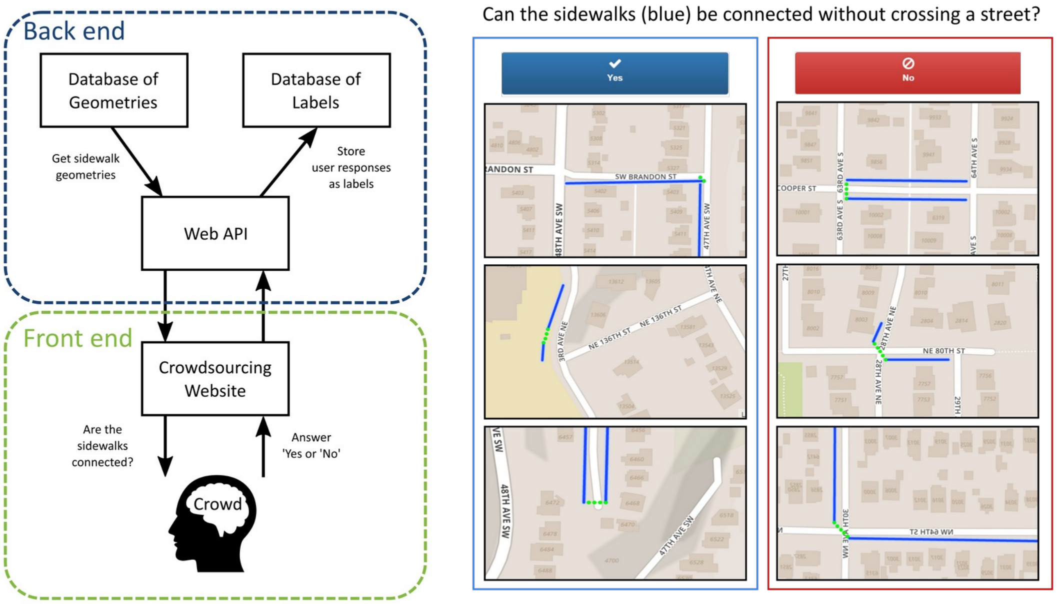


Figure 6.3: **Crowdsourcing sidewalk end pair labels**

We created a web application to crowdsource the labels necessary to apply supervised learning to the sidewalk data connection problem. The left panel shows the overall flow of information, culminating in a database of labeled sidewalk end pairs. The right panel shows label examples and shows snippets of screenshots of the user interface as it was deployed.

Ground truth data were obtained for sidewalk pairs by developing an online crowdsourcing web application through which a group of non-expert users could submit responses(Figure 6.3). The application presented users with a bird's eye view of two sidewalk segments over a basemap and asked them to decide whether the sidewalk ends should be connected (1) or not (0); we collected more than 15,000 unique labels. To validate the consistency and accuracy of non-expert responses, we compared their responses to 578 labels contributed by experts previously familiar with the dataset (i.e., AccessMap developers). We found that individual crowdsourced labels disagreed with the expert labels in fewer than 3.5% of cases.

We evaluated several common supervised learning classifiers on evenly split training and test datasets using the scikit-learn library: logistic regression, support vector machines with linear and radial basis functions, decision trees, and a naive Bayesian classifier. Every classifier generated at least a 5% improvement in accuracy over the original heuristic algorithm, with logistic regression on second order polynomial features and SVM with radial basis function producing 95.58% overall accuracy, precision over 91%, and recall greater than 84%. Spatial features derived from street and sidewalk geodata received the strongest regression coefficients, with metadata-based features trending to zero due to regularization. These results suggest that this approach may scale well to other cities. Moreover, these algorithms achieved high (>90%) accuracy with resampled training set sizes of 1000 data points, demonstrating that crowdsourced learning can be a low-cost, rapid method to address pedestrian data problems.

6.2.3 *Derivation of sidewalk networks from street centerlines: combinatorial maps and Sidewalkify*

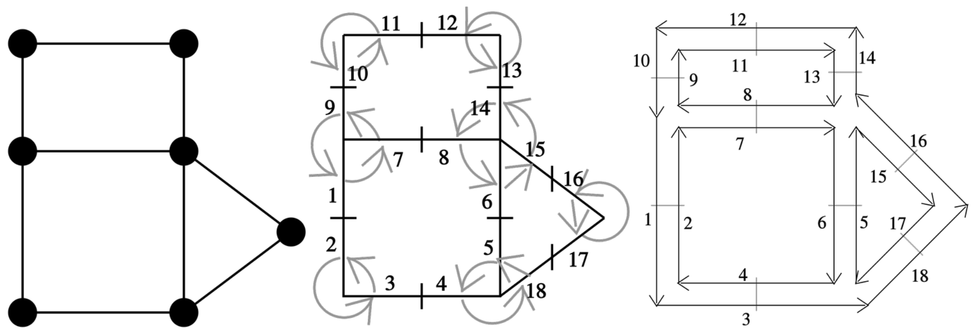


Figure 6.4: **Combinatorial Map approach to deriving sidewalk lines**

A combinatorial map approach to polyhedral extraction from an embedded graph. The approach taken by AccessMap (and implemented in Sidewalkify) uses similar logic, extracting counter-clockwise (CCW) turn paths that are then extracted into connected sidewalk paths. The left graph is a planar embedded graph, analogous to a transportation network. The center graph shows edge darts following CCW turn logic. The right graph shows the extracted polyhedral surface path, or in the case of a pedestrian network, the extracted sidewalk paths.

The aforementioned approaches to correcting connectivity errors in the Seattle DOT sidewalk dataset resulted in 90-95% accuracy in connecting sidewalk ends together correctly. However, this is actually relatively poor performance for a useful sidewalk network: if errors are distributed evenly, then the chance of a single automatically-planned trip on that network increases with the reciprocal power of the error rate. In other words, there would be a 22.6% chance that a trip covering 5 sidewalks would have an error and a 60% chance that a trip covering 10 sidewalks would have an error. This is far too high of a rate to create a practical routing tool.

After meeting with SDOT officials and probing the data more deeply, we discovered that the geometries of the sidewalks dataset published for Seattle are not separately maintained or curated as dedicated artifacts, but are derived from the streets dataset. Specifically, the sidewalks dataset is derived from running a linear offset algorithm from ArcGIS on street lines given an estimated offset distance. These geometries would extend parallel to an entire street and therefore often intersect with neighboring streets or otherwise appear inaccurate, so they are trimmed back by a constant distance to appear more correct. This strategy creates lines that are in roughly the right location, but do not retain network connectivity information and still have significant geometrical errors. It also explains the existence of the "T" intersection problem described earlier: in a "T" intersection, two streets are removed by a roughly 180 degree angle, so associated sidewalks should extend approximately the full distance, but the constant trimming creates systematically short sidewalk segments. With this understanding, it became clear that the value added by attempting to modify or otherwise "fix" these existing geometries would be less than attempting to use the source data, the street network and sidewalk offset distances, to directly infer a more accurate sidewalk network.

In previous examples of attempting to derive a sidewalk network from Seattle's municipal data, the core challenge was classification: should two sidewalk endpoints be connected or not? When returning to the street data and linear offset algorithm strategy, it became clear that there is another source of information that will more reliably answer this question: the street network itself. Given two neighboring streets that meet at an intersection, if both indicate the presence of sidewalks between them, then a connection is implied. Therefore, a more successful sidewalk network extraction algorithm can use the same linear offset strategy as employed by Seattle, but rather than trimming back sidewalks by a constant distance can instead connect sidewalks based on graph logic.

The core logic added by our approach is the use of a combinatorial map(Figure 6.4), a model for embedded graphs most often used for extracting surfaces from N-dimensional polyhedra[20]. By exploiting the sorted ordering of edges incident to nodes, darts from one edge to its topological neighbor can be deterministically derived and used to create accurate extractions of surfaces, or in this case, closed paths. The steps to redraw Seattle's sidewalks in a topologically accurate way can be summarized with the following steps. (1) The street network dataset, which is published as a set of line geometries that start and end at intersections, is inferred into a graph structure where intersections are nodes and streets are edges. (2) For every node, its neighbors are sorted by azimuth, the angle of approach. (3) A pathfinding algorithm is run that makes only counter-clockwise (CCW) turns, simulating the path that could be taken without crossing a street, assuming all streets had sidewalks. (4) Sidewalks are drawn where indicated using a linear offset algorithm. (5) Every CCW path is revisited and traversed to connect and trim sidewalks. Sidewalks that immediately follow one another along a path are connected either by trimming them to their intersection point or by connecting them with a new line segment, which is then split in half, with each half being assigned to one of the sidewalks. If a sidewalk does not have a neighbor, it should still be trimmed to match reasonably geometric expectations, which is accomplished by estimating a pseudo-sidewalk line at a constant distance from the street and trimming at the intersection point.

After running these steps for all intersections and streets in the city, the result is a topologically accurate sidewalk network with more accurate geometries. The implementation of redrawing Seattle's sidewalk network has been generalized into a software package called Sidewalkify[74], which requires only a street network with embedded sidewalk offset distances to create a topologically accurate sidewalk network. Sidewalkify can be used either as a command line interface application or as a Python library.

6.2.4 *Promising alternatives for other GIS data primitive source data*

The surveys in Chapter 5 indicate that polygonal data derived from orthogimagery are a commonly-published data source, often being available where centerlines are not. Polygonal data represents sidewalks as two-dimensional surfaces (polygons) that can be traversed, providing an opportunity to derive more specific and accurate paths and width information that would normally be available from block-long sidewalk geometries, as currently used in AccessMap. For example, in datasets where planters are represented as holes in sidewalk polygons, one could derive paths around them and estimate width changes based on such features. In combination with street furniture polygons that could represent objects such as bus stop shelters, mailboxes, or lamp posts, it would be possible to estimate a "widest-path" sidewalk line that is both more geometrically accurate and could be used to account for device constraints like wheelchair width.

Two approaches are readily available and could be adapted to derive a sidewalk network after addressing some key challenges. Straight skeletons or skeletonization algorithms are common in image processing and can also be applied to polygons, extracting an “backbone” linear representation of a polygonal feature as well as “bones” that connect it to the surrounding polygon. This algorithm would be successful for contiguous sidewalk polygons that represent the entire connected sidewalk surface, but many datasets publish polygons that have been broken up to represent different sidewalk interfaces, so these polygons would potentially need to be joined or the algorithm altered to accommodate for these issues. Another approach would be to adapt pathfinding algorithms from robotics, which derive paths in complex spaces using 2-dimensional models and line-of-sight estimations. This approach would also require joining traversable polygons and defining start and end locations or constraints that maximize the distance between the “robot” and surrounding features.

6.3 INVENTING STREET CROSSING LOCATIONS

6.3.1 *The need for guesswork*

Where a pedestrian can cross the street represents a significant technical challenge for pedestrian networks on multiple fronts. Depending on locale, including many places in the United Kingdom, pedestrians crossing the street may be legal at any location, connecting sidewalks arbitrarily along sidewalks or other footpaths. In such a scenario, there are an infinite number of potential crossing locations and even explicitly cataloging and publishing a discrete number would be potentially wasted effort: this network should be derived from the data sources themselves (e.g. street and sidewalk polygons), not manually curated. In Washington, pedestrians usually do not possess the right-of-way to cross streets, but instead must cross at intersections. In all cases, street crossing data is nearly always missing (Table 5.1), so these essential pedestrian network data must be manually gathered or inferred from scratch. AccessMap has addressed this problem using both approaches: first, by inferring likely street crossing locations algorithmically and then by collecting crossing directly in OpenStreetMap.

6.3.2 *Balancing objectives: drawing street crossing locations with Crossify*

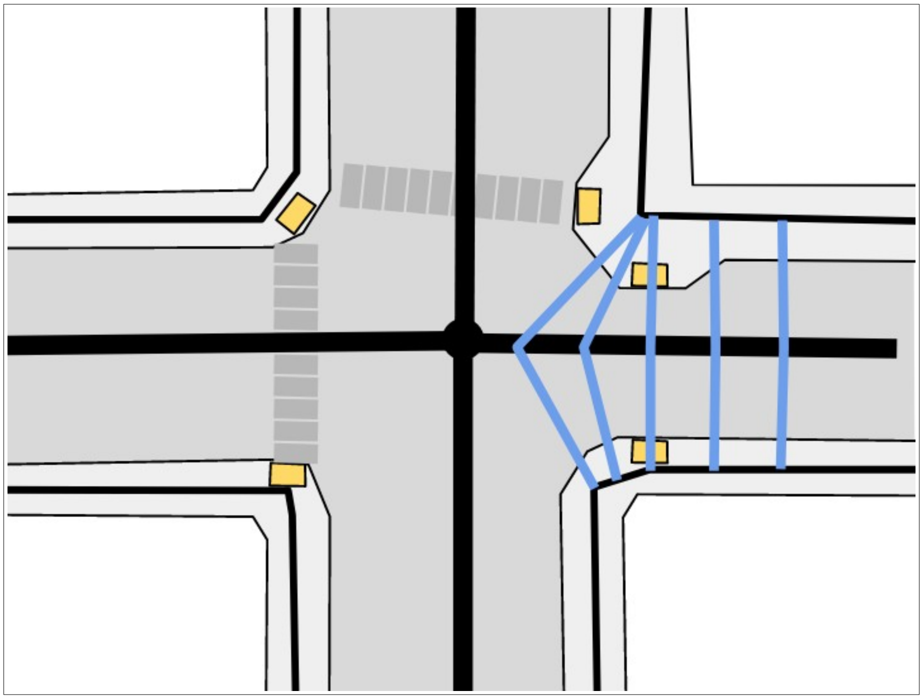


Figure 6.5: **The Crossify strategy for estimating street crossing locations.**

Crossify estimates street crossing locations by marching down a given street away from an intersection, evaluating a cost function that include distance from the intersection, length of the crossing, and the dot product of the street and crossing (proportional to angle of intersection).

Because pedestrian crossings are typically unmarked and undescribed in geodata, it was necessary to invent reasonable paths for crossing the street at intersections. For this purpose, I developed a software package entitled Crossify that will estimate street crossing lines given sidewalk and street lines as inputs. Early efforts to draw crossing locations used ideas of sidewalk "corners" to connect paths, where sidewalks had been drawn as simple lines that connected when streets met, creating hard corners. However, real sidewalk data may come from a variety of sources and more realistically match real-world variation in shape like rounded corners or wide plazas and could not use this strategy without losing wide applicability. Crossify instead attempts to balance three objectives: (1) drawing a path that is close to an intersection, satisfying a common legal requirement and pedestrian preference, (2) drawing a path that is short, capturing paths more likely to be taken by pedestrians, and (3) drawing a path that crosses the street at a roughly orthogonal angle. Crossify implements this strategy by creating a cost function that is a linear combination of these metrics, with each metric having an associated and user-definable constant, and finding a crossing that roughly minimizes the cost function. For practical purposes, this optimization is done by testing a discrete set of candidates generated by "marching" down a street away from a particular intersection, up to an include the half-way point between intersections, generating candidate crossings and the associated metrics. The crossing with the lowest cost is then retained and the process repeated for all intersections. As of writing, the user-defined constants producing reasonable crossings are 1, 0.2, and 500 for the length of the crossing geometry in meters, the distance to the intersection in meters, and the dot product of the crossing and street lines, respectively.

6.3.3 *Extrapolating street crossing locations from point data*

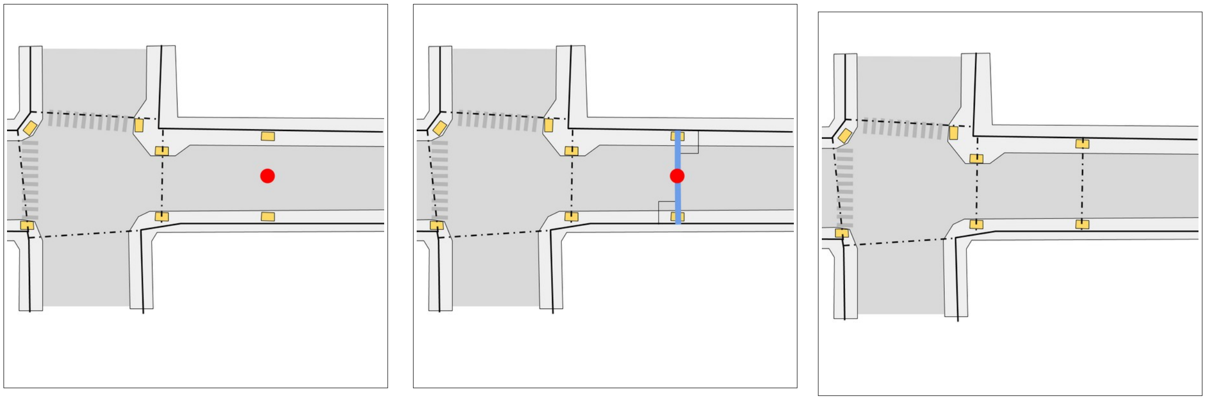


Figure 6.6: **Extrapolating crossing polylines from mid-block crosswalk point data**

The left panel shows the pedestrian network generated from sidewalk lines and street crossings at intersections as well as a point representing a mid-block crosswalk location. The middle panel demonstrates the core strategy used to process these data into additional elements of the pedestrian network: nearest points on sidewalks to either side of the street are identified via projection. The third panel shows the result of connecting these points with a new crossing line. The existing sidewalk lines are then split at the projected points to ensure graph connectivity.

While most street crossing locations at intersections can be inferred from context, there exist crossings away from street intersections, such as marked crosswalks in the middle of a block.

The Seattle DOT published a dataset of such crosswalks, collected as a point along a street. To account for these alternative crossings, we connected sidewalks on either side of each point to one another by proximity: the closest point on a sidewalk on the "left" side of a street and the closest point on the "right" side of the street (Figure 6.6).

6.4 ASSIGNING POINT METADATA TO PEDESTRIAN NETWORK FEATURES

6.4.1 *Marked crossings*

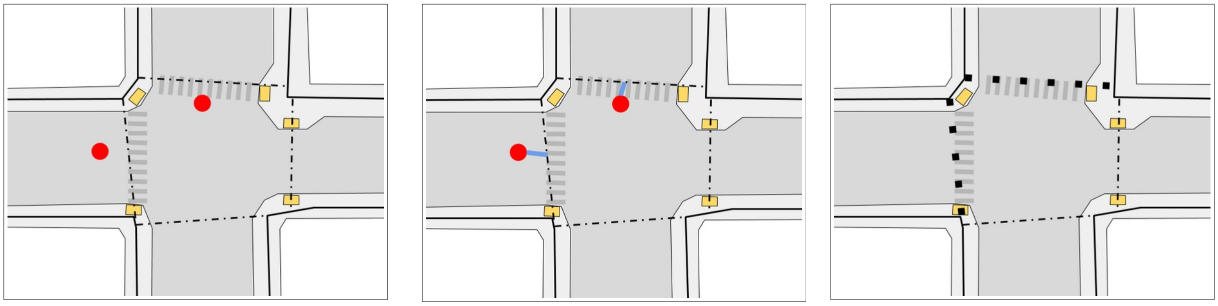


Figure 6.7: **Assigning crosswalk metadata through nearest-neighbor assignment**

Crosswalk data published as free-floating points in space can be used to annotate crossings that are already embedded in the pedestrian network through nearest-neighbor assignment. The first panel shows SDOT crosswalk data as red points. The middle panel shows nearest-neighbor assignment, where each crosswalk is assigned to its nearest neighbor and then removed from a pool of candidates. The final panel indicates that the two crossings to which crosswalk points were assigned have been updated as marked crossings.

Pedestrians prefer to cross at locations that are protected in some fashion, including traffic lights, stop signs, street markings, and curb bulbs. Therefore, a pedestrian network should include unambiguous information about these features within street crossing edges. SDOT publishes crosswalk locations (marked crossings) as point data along the street network at a location that is roughly on top of the actual markings. AccessMap Seattle uses a nearest-neighbor assignment strategy (Figure 6.7) for determining whether a crossing line is 'marked' or not: (1) a set of crosswalk points is generated, (2) for every crosswalk point in the agency dataset, the nearest crossing within a set distance is set to 'marked' and the point is removed from the set, and (3) the process is repeated until the set is empty or only contains crosswalk points that are too distant from crossings. As of writing, the crosswalk points dataset always empties, i.e. all points can be assigned to crossing lines.

6.4.2 *Curb ramps*

Curb ramps represent a similar challenge to crossings: they are ranked as important by many pedestrians but the data are usually missing or are in odd formats. Curb ramps can be found as polygons representing the actual ramp area of the curb ramp installation, as the entire curb ramp installation, as points in space roughly centered on the ramp, as points in space located at the interface between sidewalk and street surfaces, and as metadata of crossings (Table 5.1). It is not currently possible to infer curb ramps' existence by spatial relationships or statute, unlike unmarked crossings. As a result, curb data is inferred based on disparate strategies, as all three regions currently served by AccessMap have disparate data sources. For Mount Vernon, curb data is collected via the OpenSidewalks standard and is used to update crossings with "curbramps=1" based on graph traversals. For Bellingham, curb data are published as points coincident with a linear street crossings dataset that is used to update the metadata of crossings early in the data transformation pipeline: "curbramps=1" when a crossing has curb ramps on either side. For Seattle, curb ramps are coincident with sidewalk endpoints, but because those sidewalk endpoints are so inaccurate, curb ramps are later moved based on the results of Sidewalkify. Crossings are then updated with "curbramps=1" if curb ramps fall within a few meters of the crossing endpoints.

6.5 COMBINING EXTRAPOLATED NETWORK DATA WITH MANUALLY-CURATED DATA

6.5.1 *The need for manually curated data in Seattle: paths through buildings*

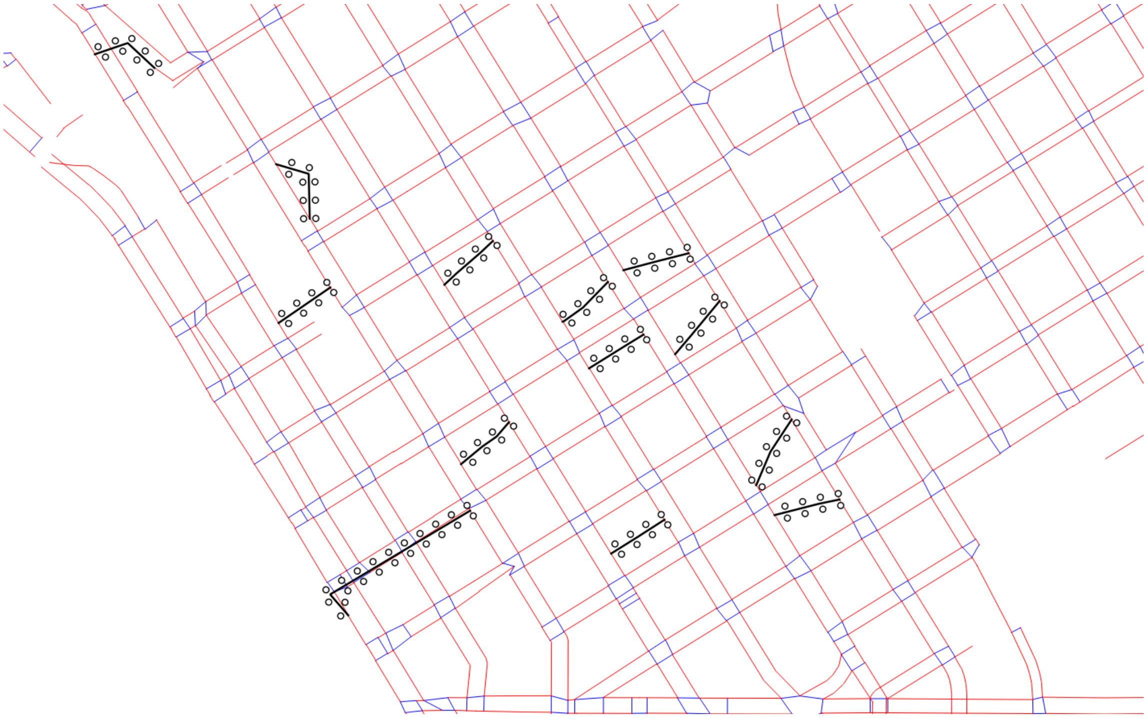


Figure 6.8: **Elevator-using paths in downtown Seattle**

Elevator-using paths in downtown Seattle (black paths with circles) were annotated within OpenStreetMap by hand. These paths were then combined with the agency-sourced AccessMap pedestrian network by nearest-neighbor snapping and splitting, similar to the mid-block crossing strategy. Elevator-using paths allow pedestrians with incline constraints to take shortcuts through elevation changes.

Downtown Seattle is adjacent to the Puget Sound and has steep inclines in the Northeast direction that often exceed 15% grade. In informal interviews, wheelchair users indicated that they shared local lore about which buildings had public elevators that could be used as shortcuts around elevation changes. For example, entering a building on a lower elevation, taking an elevator to another floor, and exiting at a higher elevation on the next block. In addition, King

County Metro, the local transit provider, publishes a manually-curated map of “accessible” paths downtown that include some in-building paths. Because downtown Seattle had been mapped in OpenStreetMap as part of a closely-related project, we manually mapped these paths through buildings using OpenStreetMap conventions, then combined these data with the sidewalk and crossing data derived from municipal sources and interpolation with Sidewalkify and Crossify (Figure 6.8). Merging logic was straightforward: identify the closest sidewalk to the in-building path, project from the building path to the sidewalk, and extend the polyline to include that path. The sidewalk was then split at this shared point to ensure graph connectivity in downstream processing steps: shared endpoints defined graph structure.

CHAPTER 7. DERIVING A PEDESTRIAN NETWORK THROUGH CROWDSOURCING

7.1 WHY CROWDSOURCING?

Access to and production of information about public pedestrian spaces raises philosophical, political, and pragmatic questions. One of the key challenges faced by this and related projects is a lack of information: despite being owned by the public and often under mandate to be collected by governmental agencies, the pedestrian network for a public space is almost never collected. Though there are often mandates to make public spaces accessible, such as ADA guidelines on public pedestrian paths, entrances to buildings, and elevators within buildings, they are often not accessible to many pedestrians. Therefore, essential information, let alone data, about whether a space can be used by any particular categorization of pedestrians is unknown and is not embedded as an expectation by the public. In addition, in informal discussions during attempts to include government agencies in data gathering efforts, officials from widespread agencies and locations indicated competing aims that stymied the gathering and publication of pedestrian network data. In short, officials and other governmental workers are enthusiastic about the importance and necessity of collecting the data, but are often constrained by concerns about opening up their agencies to legal liabilities: if the data about their pedestrian network were not just collected, but provided with a stamp of approval by the agency, it would indicate that the agency was aware of a lack of compliance and would then need to rectify the issues with a much larger budget than they currently have or open themselves to a lawsuit to that effect. While this dynamic is a consequence of complex political realities, it has the effect of limiting public discourse and our collective understanding of pedestrian spaces as well as hiding salient problems in pedestrian infrastructure. Pedestrian network data must be available for the public to understand and effectively use their infrastructure, but toxic incentives make this more difficult.

Crowdsourcing public data provides a means to work around these and other reasons that pedestrian data may be unavailable. Individuals drawing their own, public pedestrian network data could immediately make use of these data in an application like AccessMap or engage in their own analysis. Crowdsourcing therefore represents a path forward for creating a data commons independent of agencies, and may provide pressure for agencies to collaborate with such a commons, as infrastructure problems would be published and known regardless of the indications of a legal department. In addition, crowdsourcing brings up important questions of data provenance: who should control information about public spaces, make changes to it, and make use of it? A public, crowdsourced map suggests that all individuals have a right to all three.

There are potential downsides to crowdsourcing, particularly the labor requirements: if not enough individuals map a space, or update the map in that space as the infrastructure is changed, the map data could rapidly become useless. While it requires an investment of labor from the public, in our experience it is not difficult to locate enthusiastic volunteers for mapping pedestrian network information so long as it is known the data is open and used for a socially responsible cause. In addition, these data provide an incentive for agencies to work with a crowdsourced data commons, a relationship that could alleviate the burden of updating the map as infrastructure changes, placing it on agencies that already record similar information in internal (but inadequate) formats.

7.2 OPENSTREETMAP

As part of our interest in crowdsourcing pedestrian network geodata, we have identified OpenStreetMap (OSM) as a suitable sidewalk data commons. An open source, open data, user-editable map used by millions worldwide, OSM is often described as the Wikipedia of maps. Therefore, we built tools to facilitate community data validation of the sidewalk data and to simplify edits and manipulation of the data in OSM, supporting the OSM community as it maintains and contributes new, on-the-ground information. We thereby leverage the recent surge in Volunteered Geographic Information (VGI) not only for crowd-vetting (as previously proposed), but for sustainably promoting community engagement for this public-service project.

Our solution resembles ongoing, successful efforts to use OSM for humanitarian purposes (HOT). HOT has demonstrated that individuals are highly motivated to contribute geographical data for a social good. It has also shown that community engagement in mapping provides effective STEM education opportunities for students worldwide.

7.3 THE OPENSIDEWALKS PROJECT

The OpenSidewalks project encompasses all relevant efforts and pain points in crowdsourcing pedestrian network data in OpenStreetMap. Initial attempts to use OpenStreetMap to map pedestrian networks revealed several pain points and limitations we identified as missed opportunities. OpenStreetMap began as an academic and tech-forward project and maintains much of the legacy in the form of discoverability and barriers to entry for contributing and using data.

Discoverability is a key limitation of using OpenStreetMap: an enthusiastic new user will need to learn a significant amount of data primitives and idiosyncrasies in OpenStreetMap, consult disparate and hard to discover sources of information (including a mailing list with poor security practices), identify a piece of editing software appropriate for their task, and often still run into serious ambiguities. The experience of a new user is demonstrative of these limitations. If a new user decides that they want to contribute sidewalks to the OpenStreetMap data commons, they might visit the main OpenStreetMap website[82]. There, they will find a map that has been zoomed out to continent scale centered on their estimated location using their IP address. They will find a top bar with buttons titled, “Edit”, “History”, and “Export”, login and account creation buttons in the top right, and miscellaneous links between that include “GPS Traces”, “About”, and “Copyright”. If they decide to log in and click on the “Edit” button, they will be taken to the only official OpenStreetMap contribution application, the iD editor, a JavaScript map editing application. This application has a tutorial that explain how to use the tool to draw and tag (annotate) different geometrical shapes using OpenStreetMap terminology, but will not provide any resources on data standards or best practices. The new user will be placed at the main iD screen with options for clicking on existing map features or drawing points, lines or polygons that can then be annotated. They will need to discover that drawing a line and searching for the “sidewalk” preset will allow them to draw a sidewalk. To make matters worse, they will need to understand OpenStreetMap much more deeply, as there are mapping standards they will certainly violate: the lines they draw must be connected to one another to create a network, sidewalks should be connected across the street by street crossings, and street crossings should be connected to the street network by sharing a node (point). Because they must draw crossings, they should become familiar with the different crossing tags and the complex and often confusing schemas around them. In addition, annotating curb information is essential for many wheelchair users, so mapping out a space sufficiently will require learning how to apply curb tags as nodes along ways (lines). Finally, there is no way to distinguish whether a feature is missing from the map or whether nobody has attempted to audit that space for the feature, so the contributor must become intimately familiar with spaces as large as an entire city in order to conclude whether they have finished mapping out, for example, their sidewalk network.

There clearly exist a large set of barriers between an enthusiastic individual that wants to add easily-identifiable information (sidewalks in their neighborhood) to OpenStreetMap. The following sections will discuss OpenSidewalks efforts to address these limitations and more, focusing on software tooling to work around both technical and social limitations, the data schema used in OpenStreetMap at the feature level, a holistic description of mapping pedestrian spaces as connected networks, and the results of assisting the mapping of public pedestrian networks.

7.4 A MISSED OPPORTUNITY: MAPPING TOOLS

As part of the OpenSidewalks project, we hosted several “mapathons” (mapping parties) where volunteers congregate to map out a space in OpenStreetMap for a particular end. Mapathons are often a key way for local OpenStreetMap communities to socialize and organize and are often targeted as causes for social good, such as the Humanitarian OpenStreetMap Team [83], and likely serve to bridge some of the gaps in the online discoverability experience. In all, the OpenSidewalks project has hosted over 200 volunteers for mapping pedestrian spaces in the Seattle area. Preparation for and engagement within these mapathons brought serious limitations in converting volunteer enthusiasm into public geodata on pedestrian spaces: mapping tools and interfacing with data schemas.

The first barrier to contribution that presents itself to any new user is registering an account with the OpenStreetMap website, which is required for contribution. Volunteers exhibited a wide range of familiarity and competence with registering new accounts on websites and it was necessary to walk several volunteers through each step, particularly the email confirmation. In addition, significant confusion arose regarding the single-sign-on (SSO) portal offered by OpenStreetMap. Once registered, volunteers would use either the iD editor or the JOSM editor depending on the task at hand. For volunteers attempting to the JOSM editor, a Java-based "power user" desktop application for OpenStreetMap contributions, it was then necessary to set up detailed OAuth credential information within the application, which was both error-prone and required several detailed steps to implement correctly. The onboarding steps for registering a new account and having the ability to contribute to OpenStreetMap took 15-20 minutes alone. The OpenSidewalks project does not have an immediate solution to this challenge, as much of it is core functionality of the OpenStreetMap website, but has created a step-by-step manual for carrying out this process individually or in a group.

The next barrier to contribution is in translating a volunteer's knowledge and pattern recognition abilities into map data. For any sufficiently complex mapping task, OpenStreetMap contributors will need to understand the underlying data model: nodes, ways, tags, and relations. The challenges of understanding and applying these primitives correctly for mapping pedestrian spaces is covered in a later section, but as an example, mapping a linear marked street crossing requires drawing a polyline connecting to sidewalks and the street network and then applying three tags: "highway=footway", "footway=crossing", and "crossing=uncontrolled". A new user can easily identify a "marked crossing" and draw it, but accurately translating this into the correct tags (and with correct topology) is non-obvious and required 15-20 minutes of group training. The iD editor now assists with many of these challenges, but much of the core problem remains: the intent to map for a particular purpose is not matched by the tooling and consequently requires significant training and familiarity with core OpenStreetMap data primitives and tagging standards. To assist this, OpenSidewalks has created manuals and guides for topologically correct mapping of pedestrian networks in OpenStreetMap, as the exact means by which to combine tags and geometries is non-obvious even after combining all official OSM resources on the topic.

Finally, there are several pieces of data that are not well-suited to the aforementioned "mapathon" format, where volunteers "armchair map" (remote mapping from imagery) as they are not clearly visible from satellite photographs or alternative imagery sources (such as Bing StreetSide™) are too out-of-date. In these cases, an in-person survey is necessary to extract this information that often includes essential pedestrian network data like footpaths or curb ramps under trees or tunnels. To assist in these tasks, the OpenSidewalks project has generated two approaches: adapting the StreetComplete[70] Android application for on-site OSM contributions and an imagery gathering strategy based on the Mapillary[59] service.

The StreetComplete Android application gamifies the on-site contribution of data to OpenStreetMap by identifying existing OpenStreetMap features for which data is missing and turning that lack of information into a targeted quest. For example, street lines frequently lack surface type information (gravel, asphalt, etc), so StreetComplete has a street surface quest where participants can tap on a quest bubble on the map and answer a human-readable question that fills in and contributes appropriate OSM tags. This approach successfully bridges much of the enthusiasm-contribution gap identified earlier, so we adapted this application to create a prototype “GoInfo” application with new and prioritized quests for accessibility and pedestrian-relevant information, including properties of bus stops that assist blind pedestrians with navigation (presence of shelters, garbage cans, benches). GoInfo was successful at mapping out these features in testing, but was limited by the necessity of having existing features to annotate (the primary challenge remains the collection of a minimal pedestrian network) and because even with existing features, querying for missing data is a complex network problem that will require significant changes to the internal quest structure, such as when curb information is missing. For the last example, one would need to phrase this question in order to identify missing curb data: at the node where crossings and sidewalks meet, but not at a node internal to a sidewalk or a node shared by two or more sidewalks, check for a lack of the “kerb=*” tag. This query could easily be asked within an internal network representation, but StreetComplete is based on receiving quests directly from Overpass API[84] queries, for which this question is relatively complex.

The use of the Mapillary service for surveying was chosen as an alternative to on-site direct contributions of data, and amounts to the creation of a crowdsourced version of Google Street View™. With Mapillary, individuals contribute geotagged photos with manually-adjusted locations and headings and the service provides an intuitive interface for investigating them, including stitching into larger scenes. One advantage of this approach over on-site surveys is that the photographs can be used to complex spaces from an off-site location where tools better-suited for complex mapping are available. In addition, photos on Mapillary can be used to map new features in later mapathons without the need for an in-person survey. OpenSidewalks formalized a workflow for collecting and making use of these data for mapping pedestrian spaces as part of their mapathons. In the future, computer vision offers another means by which to exploit geotagged photographs, with the potential to extract pedestrian network data directly from imagery without manual human curation. Such “passive” data collection efforts have a low bar to entry and are a likely path forward for more rapid crowdsourcing of public geodata.

7.5 SCHEMA CHALLENGES IN A RADICALLY-OPEN DATABASE

OpenStreetMap’s open structure and community philosophies suggest a “globally local” approach that raises data federation and balancing of stakeholders concerns to the forefront. On a technical level, as a fully-open database, any user can create any data they want with the data primitives offered, including arbitrary key-value tag pairs. Standards are enforced through bans for vandalism and community oversight: more-active OpenStreetMap users identify issues and fix them and often reach out to mappers to try and resolve potential conflicts. The process of standardization is varied and not always clear, however. There is no official community process by which a mapping strategy is ordained as a standard, but there are sources of standardization that are often enforced through various means. First, the OpenStreetMap community hosts a wiki for community discussion and guides for mapping, including schema definitions for tags and their combinations with geospatial data primitives. This wiki also implements a proposal process by which new “tagging” standards are enshrined in wiki pages, requiring a certain proportion of votes for or against in order to become adopted. However, participation in this proposal process is often low, with only 10-20 voting participants for standards with a worldwide impact, and there is often disagreement with adopted standards between stakeholders. Furthermore, another source of standardization often holds greater sway or otherwise leads to disagreement: editing software. OpenStreetMap editors (software) often create presets or otherwise hide the tags used from users, improving the user experience, and such presets represent an executive decision by the software developers on behalf of users. If software developers disagree with the OSM wiki, mailing list, or any other OSM platform or user, this can lead to conflicts in standardization and questions about how standardization should occur. For example, the “subtags” for crossings exhibit significant limitations that will be discussed in a later section, and attempts to address them by the authors of the iD editor have often disagreed with the OSM Wiki. Prior to establishing a preset, “crossing=*” subtags would often differ based on locale even when describing the same feature: a marked crossing. Some would use the “crossing=zebra” tag and others would use the “crossing=uncontrolled” tag to mean the same thing, with neither adequately capturing the idea of a marked crossing. The iD editor initially reflected the tag more commonly used outside of Europe, “crossing=zebra”, a point of frequent contention. It then adopted the “crossing=marked” convention, which more directly describes the reality of the feature on the ground, but because this disagrees with the OSM wiki it has also coincided with community disagreements. Thus, the power over standardization is multifacted, including community members on a variety of fora, online and offline, and the creators of OpenStreetMap editors / other systematic contribution tools. In addition, because editing is done by both volunteers and individuals paid by corporations to map specific details, interests in tagging standards do not always align. Despite this, OpenStreetMap has managed to establish and maintain fairly regular standards for street networks and related features, but is still a hotbed for implementing a sufficiently descriptive sidewalk network.

7.6 MANAGING COMPLEXITY: TIERED MAPPING AND AUDITS

One of the primary challenges of crowdsourcing pedestrian data is that the number of distinct features to survey is more than can be annotated by a single individual in a reasonable amount of time. For example, an expert in the OpenSidewalks schema will take 30 seconds to map the crossings and curb information at a single intersection and must cognitively balance 5-6 different rules simultaneously to ensure correct mapping. We therefore advocate for a tiered approach, whereby an annotated pedestrian network is created in a hierarchical fashion. The first tier is to annotate the locations essential pathways with minimal annotations: sidewalks, street crossings, whether a street crossing is marked, curb ramps/curbs, and paths through curb ramps.

The need for sequential mapping of spaces for different attributes raised another important challenge: auditing the completeness of mapping. Because OpenStreetMap offers no tracking tools for audits nor recency of audits, there is no way to distinguish between map features that are missing because they do not exist vs. map features that are missing because no mappers have surveyed the area for a particular feature. In cases where these questions cannot be resolved by expected but missing tags (e.g., “surface=*” in StreetComplete), it is necessary to create tooling that facilitates the targeted mapping of features, including pedestrian network features. Many projects for managing these concerns exist in OpenStreetMap, including the Humanitarian OpenStreetMap Team Tasking Manager[49] and Field Papers[85], which split up a space into smaller tasks with instructions on what to map. The OpenSidewalks project has made use of the

OSM Tasking Manager to import data, but none are suited for structured, tiered edits, particularly for pedestrian networks. This is illustrated by the example of separately mapping sidewalks and crossings. To avoid breaking routing engines and inviting the ire of the OpenStreetMap community, sidewalks should not be mapped without being tied to the street network by street crossings. Therefore, if it is desirable to separate the tasks of street crossings and sidewalks, street crossings should be mapped first, followed by neighboring sidewalks. Specifically, all crossings associated with a sidewalk segment should be mapped before the sidewalk segment itself. While the OSM Tasking Manager could be leveraged to create separate tasks for crossings and sidewalks, there is no mechanism by which to enforce the order of mapping, let alone guide the holistic simultaneous mapping of both: once all associated crossings are mapped, a sidewalk task should become available. Such structured edits will likely be important for converting enthusiasm among volunteers into productive mapping and are left for future work.

7.7 Mixing strategies: importing agency data to OpenStreetMap

OpenStreetMap has clear benefits as a data commons, a home for exchanging data on pedestrian networks. One such exchange could potentially be from governmental agencies (such as SDOT) and the OpenStreetMap community. To attempt to jump-start the mapping of the Seattle pedestrian network and explore this space, the OpenSidewalks team hosted several mapathons focused on reviewed “imports” of AccessMap sidewalk and crossing data. This process was facilitated by a modified, self-hosted version of the OSM Tasking Manager, on which AccessMap data split by the street network were hosted as distinct tasks, representing around-the-block polygons. Volunteers would identify a “task” area to claim for themselves, which included instructions for mapping and checking the data to import as well as a link to the staged import data in .osm XML format. They would then open this data within the JOSM editor (the iD editor does not have import functionality) and “conflate”: ensure that the imported data was accurate and appropriately contextualized within the map, then submit directly to OpenStreetMap. This process was used to successfully map the University District neighborhood of Seattle, Washington in the free time of volunteers over a period of a week.

The experience of “importing” this neighborhood brought up concerns and future directions for such work. Foremost, it took approximately the same amount of time to “import” these data as it would take to map them from hand. Therefore, it may be counterproductive to use import strategies without improving the efficiency of tooling (pre-automating some of the mapping tasks) or to ensure that there is additional added value within the imported data such as surface information or a foreign key to an agency that will continue contributing after the import. In addition, this import involved adding crossings and sidewalks simultaneously, constituting too large of a task for new contributors, who would take five to ten minutes to map a single task, and suggesting the need for tiered, structured editing as detailed in the previous section. Finally, our approach was still relatively manual and depended heavily on onboarding and teaching of conventions: custom-made tools and workflows for non-experts will be necessary for getting the most out of volunteer efforts.

7.8 DATA CONTRIBUTIONS RELATED TO OPENSIDEWALKS

7.8.1 *The University of Washington*

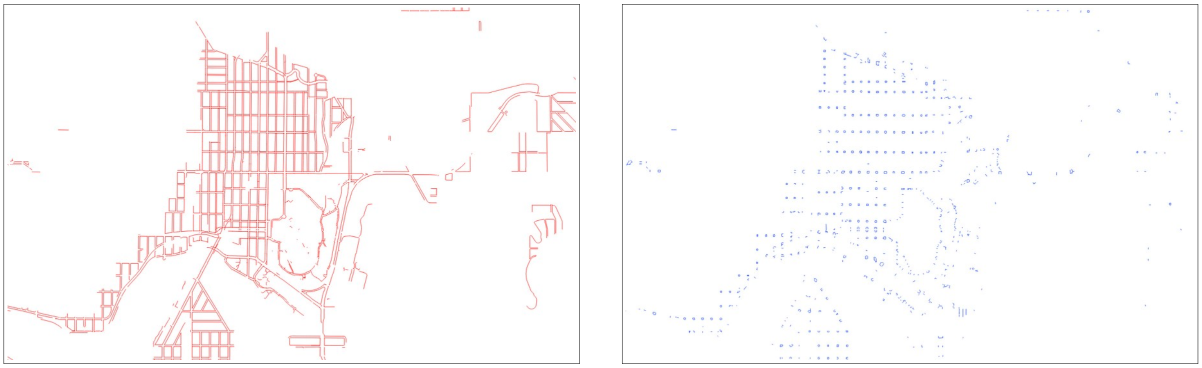


Figure 7.1: **The University of Washington OpenStreetMap pedestrian network**

The University of Washington has a highly-detailed pedestrian network in OpenStreetMap, with thorough mapping of sidewalks, campus footpaths, stairs, street crossings, and curb ramps. Sidewalks are red, crossings are blue. These data represent approximately 1500 sidewalk lines and 1000 crossing lines.

AccessMap was primarily developed at the University of Washington. Consequently, many of the mapping parties and testing efforts at developing a coherent and valid pedestrian network schema were focused on mapping the University of Washington campus. As a result, the

University of Washington campus has a highly-detailed pedestrian network that includes sidewalks, campus footpaths, stairs, street crossings, curb ramps, building entrances, pedestrian areas, pedestrian streets, and connections to transit stops.

7.8.2 *The Seattle and Bellevue, WA Areas*

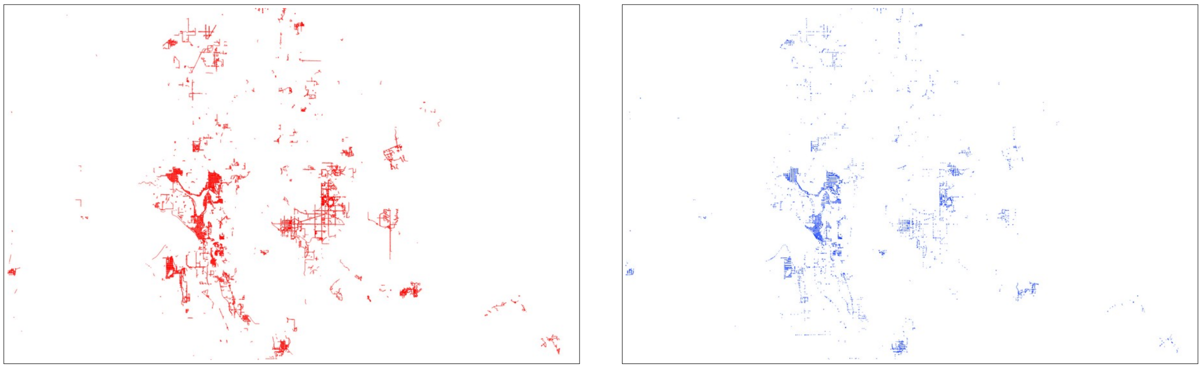


Figure 7.2: **The greater Seattle area OpenStreetMap pedestrian network**

Large segments of Seattle, WA and Bellevue, WA have been mapped according to the OpenSidewalks schema. Seattle serves as a testbed for detailed pedestrian mapping in OpenStreetMap and has been the site of targeted tests of import strategies of AccessMap data. The majority of Bellevue has been mapped directly as part of coordinated Tasking Manager efforts by students with expert oversight and validation. Sidewalks are red, crossings are blue. These data represent approximately 15,000 sidewalk lines and 8000 crossing lines.

The Seattle area has served as a similar testbed for detailed mapping according to the OpenSidewalks standard as well as a testing location for hybrid import strategies, where AccessMap data is staged for import into OpenStreetMap and then manually reviewed for accuracy. This hybrid strategy led to the complete mapping of the University District neighborhood surrounding the University of Washington. Testing out alternative "power user" approaches in JOSM led to the mapping of several surrounding neighborhoods as well, including the Eastlake, Downtown, Belltown, and large sections of Montlake, Wallingford, South Lake Union, and Ballard neighborhoods. Finally, the Judkins Park neighborhood was mapped and evaluated in partnership with Sound Transit in order to evaluate the pedestrian conditions surrounding a planned light rail station, and has been mapped with great detail.

Bellevue, WA was mapped as part of efforts to demonstrate the feasibility of student mapping. High school students were onboarded using OpenSidewalks documentation and presentational materials and then allowed to map downtown Bellevue with expert supervision. Minimal expert intervention was required after a few hours of observation. A final review was conducted by OpenSidewalks expert mappers and few changes were made. As a consequence, the majority of Bellevue is mapped according to the OpenSidewalks standard.

7.8.3 *Mount Vernon, WA*

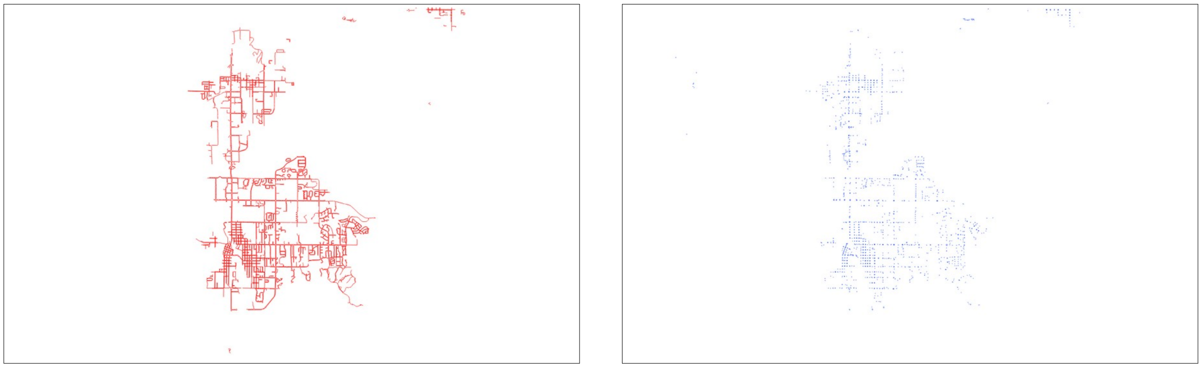


Figure 7.3: **The Mount Vernon OpenStreetMap pedestrian network**

Mount Vernon, WA has been thoroughly mapped by close collaborator and a member of the OpenStreetMap US leadership, Clifford Snow, leading to its integration into AccessMap. Sidewalks are red, crossings are blue. These data represent approximately 2000 sidewalk lines and 1500 crossing lines.

Clifford Snow was an early collaborator with the OpenSidewalks team, helping with communicating between various stakeholders, including ourselves, and bridging gaps in understanding. In agreement with the need to map the pedestrian network, he mapped the entire city of Mount Vernon, Washington according to the OpenSidewalks standard, leading to the first purely OpenStreetMap-based implementation of AccessMap. Mount Vernon therefore represents an example of an OSM-expert-mapped case study for the tight integration of pedestrian and non-pedestrian transportation networks as well as non-sidewalk, non-crossing footpaths.

7.8.4 *San Jose, CA*

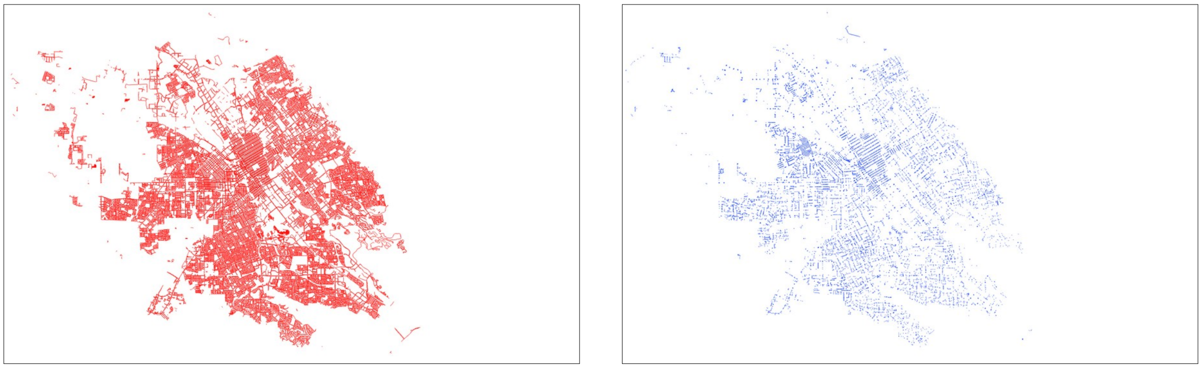


Figure 7.4: **The San Jose OpenStreetMap pedestrian network**

The San Jose pedestrian network in OpenStreetMap. Sidewalks are red, crossings are blue. These data represent approximately 16,000 sidewalk lines and 17,000 crossing lines.

San Jose was mapped according to the OpenSidewalks standard primarily through the efforts of two local members of the OSM community (Vivek Bansal and Minh Nguyen), one with an association with the local transit agency. All sidewalks and crossings in the region were mapped and expert reviews show high accuracy and detail of all spaces. Curb interfaces were not mapped, preventing the immediate integration into AccessMap, but crossings were mapped in such a way that the labeling of curbs and curb ramps can be easily split into easily-answered, easily-crowdsourced questions.

7.8.5 *Austin, TX*

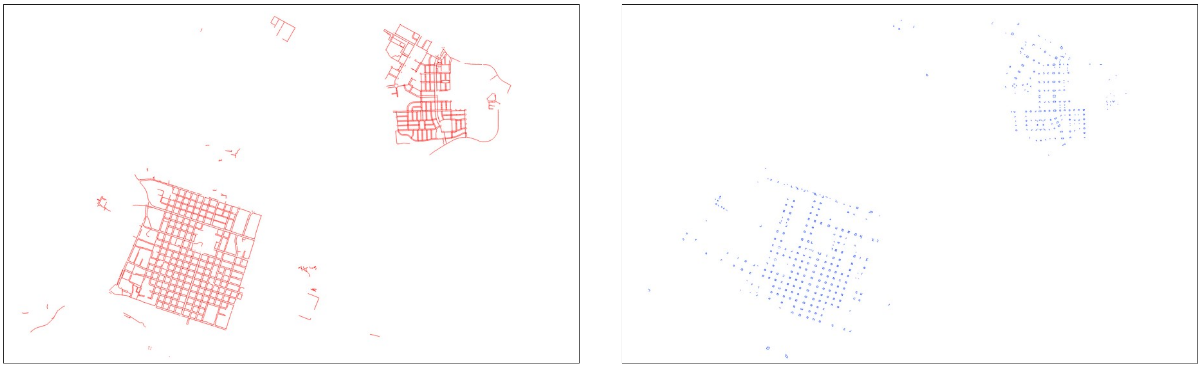


Figure 7.5: **The Austin OpenStreetMap pedestrian network**

Austin has been heavily mapped using a hybrid data approach, with an intern making modifications exclusively through the OpenStreetMap iD editor. Sidewalks are red, crossings are blue. These data represent approximately 3000 sidewalk lines and 1200 crossing lines.

In collaboration with the local transit authority, the OpenSidewalks team at the Taskar Center for Accessible Technology has mapped out large sections of Austin, TX, using this opportunity to test new pedestrian network features types.

CHAPTER 8. INCLINE (AND ELEVATION) ESTIMATION FOR PEDESTRIAN NETWORKS

8.1.1 *Background and motivation*

Steep hills are frequently cited as a concern by many pedestrians regardless of mobility device[51] and incline information has been incorporated into many maps as elevation profiles (Google Maps™, Bing Maps™, Apple Maps™) or even incorporated into routing engines (OSRM[86], OpenTripPlanner[46], GraphHopper[40], OpenRouteService[77]). Local, direct measurements of inclines are nearly always absent, so these engines make use of digital elevation models (DEMs): raster data where pixel values correspond to estimated elevations. However, the utility of these integrations for pedestrians is frequently hampered by the resolution of the data sources.

8.1.2 *Existing approaches and their limitations*

Shuttle Radar Topography Mission (SRTM) DEMs offer nearly worldwide scope with seamless coverage for all population centers, but are limited to one arc-second (roughly 30 meters at the equator) resolution. Some tools, such as OSRM or OpenRouteService, will automatically integrate this data in order to create estimated inclines and elevation profiles. More specifically, DEM data is integrated as inverse-distance-weighted (IDW) estimations of heights at specific network locations, sometimes with interpolation through tunnels. Due to the large ratio between the resolution of SRTM DEMs and the interpolated elevations along the pedestrian network, it is difficult to rely on these estimates when producing elevation profiles. In addition, such as has been the case for OpenTripPlanner, bridges are often poorly-modeled, where the DEM has eliminated the bridge from consideration entirely, so the elevation profile of traversing a bridge actually matches a trip that traversed the river below.

The National Elevation Dataset (NED) is a seamless DEM collected for the lower 48 states of the United States, with 1/3 arc-second (roughly 10-meter) resolution for all covered areas and can also be incorporated into such trip planners. However, use of the NED reveals new challenges with low-resolution data: the influence of buildings on estimated bare-earth elevations. In reality, DEMs such as the SRTM and NED are derived from source data that includes terrain (DTMs) like bridges and buildings and are algorithmically flattened in raster processing algorithms. These approaches do not actually completely remove the influence of buildings or bridges, but simply removed most of their influence. This does not pose a significant of an issue for street traffic routing applications because elevations are calculated at intersections, well away from large buildings, but for pedestrian networks, important intersections may be only two or three meters from a building.

8.1.3 *Strategies employed by AccessMap and their limitations*

Early iterations of AccessMap attempted to use an IDW or bilateral interpolation strategy for estimating elevations at points along the pedestrian network and created clearly erroneous information, suggesting that very steep hills were relatively flat, that relatively flat areas were fairly steep, and that inclines changed significantly over the length of a block even though they were relatively smooth and consistent in reality. The culprit was the extent to which building DEM pixels were picked up by the interpolation strategies.

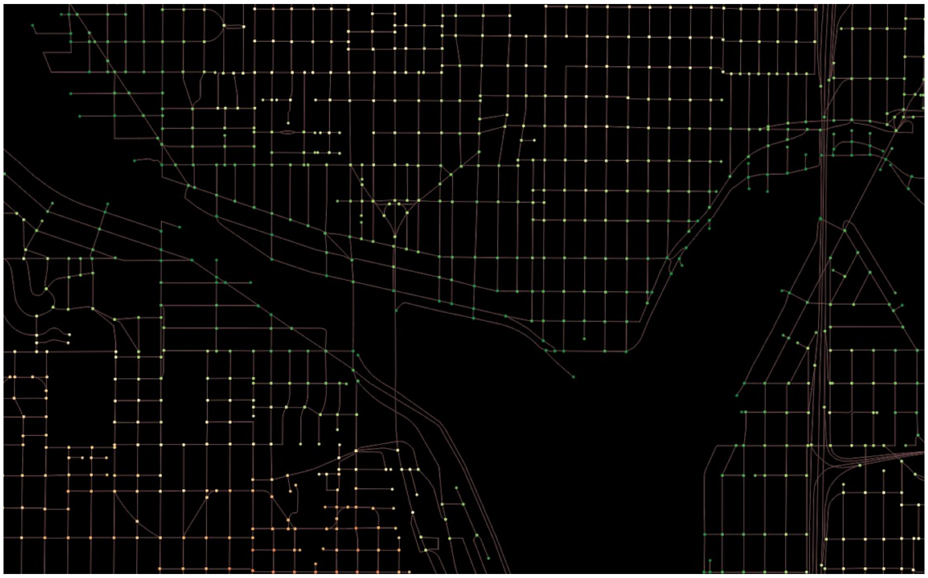


Figure 8.1: **Intersection-based elevations**

Early versions of AccessMap relied on elevations estimated at street intersections to improve incline estimates due to the interference of buildings on the underlying DEM. These point data were used to create a TIN model from which interpolated elevations could be estimated.

The incline estimation used by AccessMap then moved to borrow the technique used for streets: only elevations at intersections were estimated, created a set of elevation points distributed throughout the urban space (Figure 8.1). Then, a triangulated irregular network (TIN) was created over this point data and used to interpolate elevations along the pedestrian network. This strategy was effective for sidewalks, likely due to their similarity to street geometries and because averaging over longer distances leads to better estimations. However, incline estimates for street crossings were not accurate and had to be removed and ignored, as these (usually relatively flat) spaces had no data that would communicate a flatter situation: only intersection-to-intersection height differences were represented.

Figure 8.2: **Mask-based elevation estimation**

The current version of AccessMap identifies regions where the underlying DEM data cannot be trusted due to overlapping with buildings, bridges, or tunnels. White regions represent the masked-out pixels while colored regions indicate elevation, with yellow high and blue low.

The current iteration of AccessMap makes use of a masked NED DEM strategy (Figure 8.2). In this strategy, regions of expected low accuracy are explicitly masked out of the dataset: buildings outlines and bridges are exported from OpenStreetMap, buffered by the resolution of the dataset (10 meters), and intersected with the DEM. Overlapping pixels are masked out, leaving only data missing buildings (assuming good mapping within OpenStreetMap). The current strategy then estimates as many points as possible along the pedestrian network and interpolates the rest, still excluding street crossings. Future work could return to the original source digital terrain model (DTM) data and extract likely buildings / bridges, etc and use these to create the mask, reducing reliance on manual mapping efforts in OpenStreetMap.

8.1.4 *Future directions*

Lidar-based elevation DEMs are increasingly available and provide much higher-resolution information, which would directly address several of the limitations requiring such extreme engineering to produce AccessMap. Such data is often available at sub-meter resolution and can distinguish curbs from sidewalks from buildings, offering additional opportunities to estimate sidewalk networks in combination with other data sources, such as polygonal data. However, all overhead imagery-based estimations of elevation are taken from a limited set of overhead views and can therefore lack data for important regions. For example, production of DEMs from Lidar data requires the automated removal of foliage (trees, e.g.), filling in the surface elevation spaces with interpolation; the actual elevations underneath covering features like trees or bridges are often unknown. In addition, many pedestrian spaces are covered from above like tunnels, sky bridges, paths through buildings, or amphitheaters. In such cases, direct measurement may be more appropriate in future work using means including photogrammetry techniques from ordered photography or video (drone or smartphone) or using a smartphone’s magnetometer, accelerometer, and gyroscope to directly estimate the direction and magnitude of a surface’s incline.

CHAPTER 9. TOWARDS A PEDESTRIAN NETWORK DATA STANDARD

9.1 THE OPENSIDEWALKS SCHEMA

Inadequacies in standardization efforts and sufficiently-descriptive tags pose a significant challenge in using OpenStreetMap for crowdsourcing pedestrian networks. As part of addressing these concerns, we have developed an initial schema definition, the OpenSidewalks schema[75], to provide an internal standard for use with AccessMap that maps directly from OpenStreetMap tags, with OSM-compatible extensions. This schema serves as a target standard for pedestrian network data from disparate sources (all three cities currently served by AccessMap, each with unique data sources, are transformed into the OpenSidewalks schema format) and to act as a testing ground for the creation and proposal of OSM tagging standards.

The core of the OpenSidewalks schema is derived from the OpenMapTiles schema[28], itself a close mapping to the OpenStreetMap schemas described on the OSM wiki. Therefore, the core pedestrian network attributes described in the OpenSidewalks schema derive from OpenStreetMap tags in a one-way flow: they either match OSM tags precisely or place mutually exclusive tags under a single namespace, such as tunnels vs. bridges. To maintain consistency with the goal of improving OpenStreetMap tagging, additions made to the schema derive directly from existing tags on the OpenStreetMap wiki or represent schema additions and alterations deemed necessary by the OpenSidewalks team.

Elements of the schema that extend on OpenMapTiles while mapping 1:1 from current OpenStreetMap standards include curb (kerb) tags, surface tags, a pedestrian access override (“foot=yes”), opening hours, path width, z-index tags, and service streets (driveways, parking lots). Elements of the schema that represent schema changes or additions include the street crossings schema (“marked” vs. “unmarked”), a currently unused ramp tag, and an incline tag that represents specific incline estimates as a grade.

The current OpenSidewalks schema is not solidified, but represents a path forward for standardizing pedestrian network data within and without OpenStreetMap. An example of the schema can be seen in Figure 9.1.

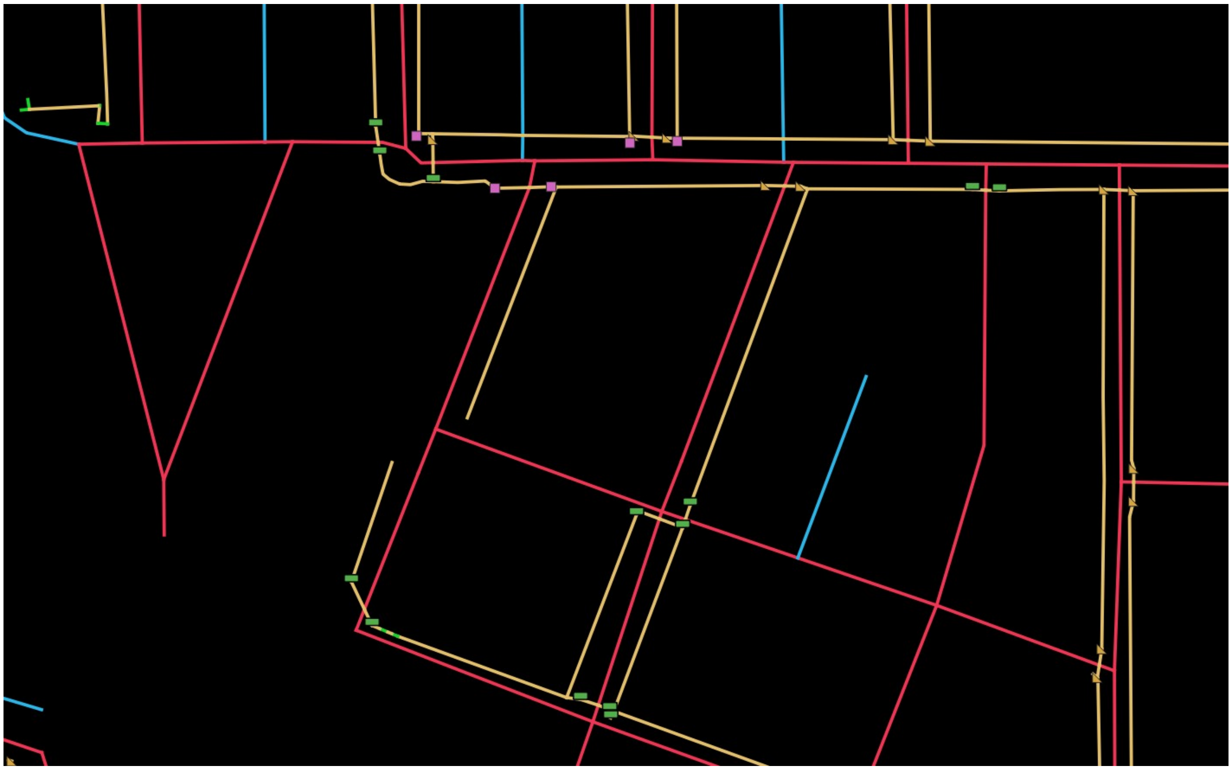


Figure 9.1: **The OpenSidewalks schema implemented in Mount Vernon, WA**

The OpenSidewalks schema as implemented in Mount Vernon, WA. All data are transformed from OpenStreetMap data into a holistically mapped pedestrian network. The OpenSidewalks schema supports fallback on the street network, so streets (red) and alleys (blue) are included. Sidewalks (light orange) and crossings (yellow) constitute the primary network and are broken up by segments of stairs (green, lower left) and conditional barriers like curbs (green are flush, purple are raised, orange are curb ramps). All intersecting paths are networked at the intersection point.

9.2 HOLISTIC MAPPING OF PEDESTRIAN NETWORKS IN OPENSTREETMAP

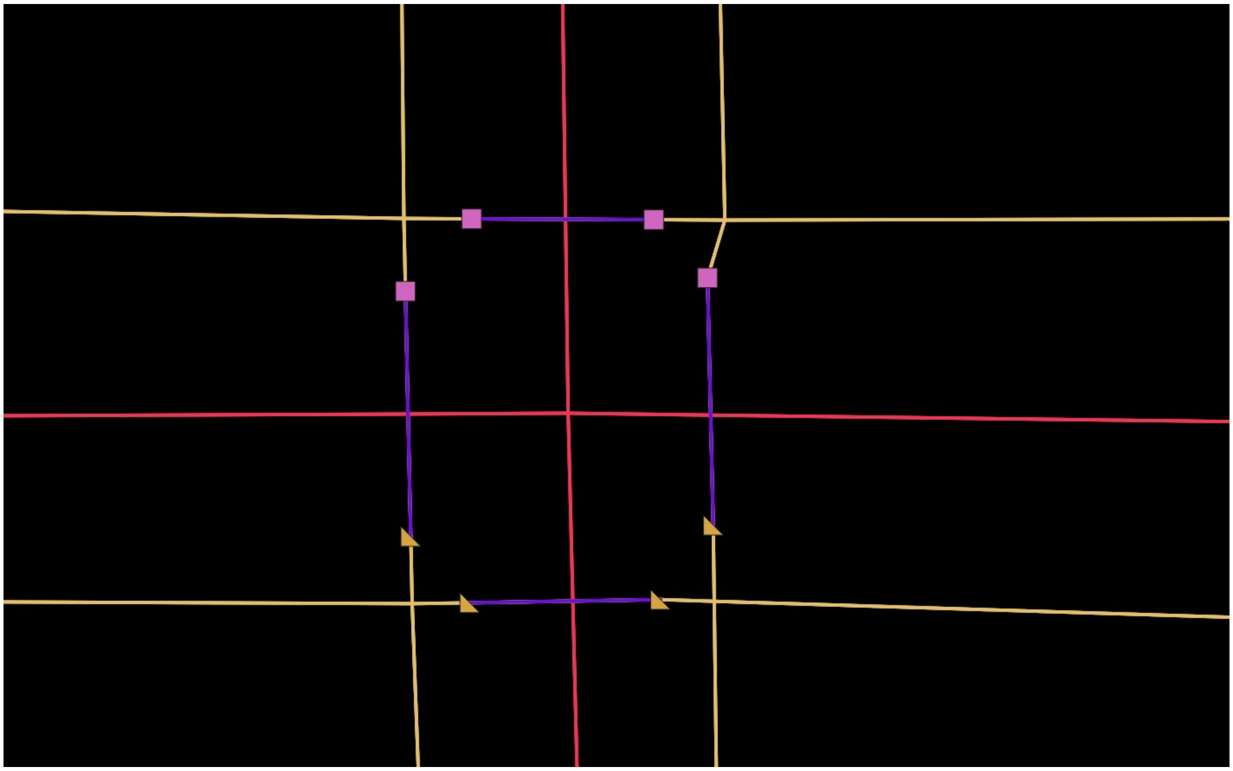


Figure 9.2: **Holistic mapping with the OpenSidewalks standard**

Defining feature-level schemas is insufficient to specify how a pedestrian network should be mapped. The OpenSidewalks schema describes exactly how schema elements should interact. Crossings (purple) should be joined to curb information at either end (raised curbs are squared, lowered are triangles). These curbs should be separated from primary pathways like sidewalks (primarily yellow lines) by short footpaths that have yet to be strictly defined (secondary yellow lines), representing the path taken from sidewalk to curb. Crossings should intersect with and share a node with streets (red), ensuring connectivity between the street and pedestrian networks.

Both the OpenStreetMap schema definitions as well as the core of the OpenSidewalks schema definition are focused on how to annotate geospatial networks one unit at a time: a single street, a single crossing, or properties of each. However, real-world mapping of networks requires an understanding of how these elements should be connected, in what order, and exactly where annotations should be placed relative to these network descriptions (Figure 9.2).

For example, a raised curb (kerb) in OpenStreetMap is defined by the “kerb=raised” tag and when one is encountered along a path, it is added as a node within that path. In the case of a curb ramp, where should this node be placed? Many mappers will place this node along the sidewalk centerline where a crossing meets the sidewalk, others will place it fully within the crossing line, and some will create an additional non-sidewalk, non-crossing footway representing the path from a sidewalk centerline to the curb and place the curb tag at the node shared with the crossing way. The OpenSidewalks schema therefore also include a more holistic description of modeling pedestrian network features in combination, specifically recommending the last strategy: curb data on nodes should be added to a node shared by a crossing and a distinct path from the sidewalk to the curb. The specific justification for this decision is based on meeting three requirements: (1) removal of ambiguity, particularly when routing, (2) data should be annotated where it physically exists on the map, and (3) missing information should be identifiable through missing tags or network traversals. Regarding ambiguity, the strategy placing a “kerb=raised” node along a sidewalk centerline would imply a falsehood: that there is an actual raised curb barrier if a pedestrian continued along the sidewalk and did not cross the street. Placing the curb node away from the sidewalk also allows for the node to be placed at the exact curb interface location, making the association clear. Finally, if kerb nodes are to be added internally to crossings, searching for instances where these data are missing becomes difficult: crossings may be split into multiple segments and identifying the locations where a curb might be expected, but is missing, is complex and extremely difficult. In contrast, if curb data is to be expected on the node where a footway and crossing meet end-to-end, this can be identified easily through network traversals and staged for a StreetComplete-like task.

Similar challenges arise for other holistic mapping scenarios, including how many curb interfaces to annotate when there are far-offset curb ramps (a pedestrian not preferring a curb ramp would not use it, in that case), the need for crossings to share a node with streets they cross, when to use the “footway=sidewalk” tag vs. not, and how to represent connections between stairs and other footpaths, with the OpenSidewalks schema providing guides and figures for these scenarios.

CHAPTER 10. THE CHALLENGE OF INDIVIDUALIZED PEDESTRIAN ROUTING

10.1 BORROWING FROM SHORTEST-PATH AUTOMOBILE ROUTING: USING DIJKSTRA'S ALGORITHM

10.1.1 *The form and function of an individually-parameterized pedestrian cost function*

An approach to efficient and guaranteed optimal graph shortest-path finding was developed by Edsger W. Dijkstra in 1956. Dijkstra's method is widely deployed for finding shortest paths for a wide range of domains, framing the "shortest" path as one in which the sum of path weights is smallest over all possible paths from and to any given node (A to B routing). Path weights may be physical distance in the case of transportation networks, in which case a shortest path is literal, or they can be any arbitrary metric or induced value, such as hard-coded to 1 (the shortest path is then the one with the fewest edges) or a time estimate (essentially how sophisticated maps like Google Maps™ find best paths for car traffic). Importantly, the weights are calculated independently and are usually even precalculated: the algorithm achieves efficiency and optimality by assuming that weights do not change during the traversal. This last condition may not hold as well for pedestrians as it does for cars due to real-world conditions like fatigue, but shortest-path algorithms like or derived from Dijkstra's algorithm are a clear path forward for probing what it means to discover the "best path" for a particular pedestrian.

Whereas a street network is relatively simple and the concerns of cars are roughly uniform and can be formulated in terms of simple metrics like speed or the use of major highways, pedestrians exhibit wide variation in their preferences and abilities to navigate urban pedestrian environments. Therefore, while car traffic is amenable to precalculation of weights and advanced graph traversal strategies and optimizations, inclusive pedestrian routing must begin with a cost function that weighs a complex set of concerns that can be individually parameterized. For example, a particular pedestrian may have an uphill maximum incline of 8.5%, a desire to only use intersections with curb ramps, an interest in using elevators through buildings when available, and trouble navigating on gravel. This represents preferences regarding four distinct feature types, three of which can have two answers and one of which, assuming incline should be discretized to half percents, has at least 30. There are therefore 240 combinations regarding just these four features, and it would therefore necessitate 240 distinct weights precalculated on the graph in order to account for pedestrian diversity. While enumerating distinct weights or partial weights may eventually be an appropriate approach for pedestrian routing, diversity and flexibility are the current key elements of exploring the problem of representing pedestrian navigation.

Because precalculating weights is not reasonable at this time, shortest-path pedestrian routing should take the form of a dynamic cost function that generates weights on-the-fly during the graph traversal step of a shortest-path algorithm. Because traversability depends on individual preferences, this cost function must be parameterized by user inputs at runtime.

$$\text{cost} = \alpha_1 g^4 + \alpha_2 r + \alpha_3 l + \alpha_4 c$$

g is grade (incline))
r is presence of a curb ramp
l is length of path
c is path is a street crossing

Figure 10.1: **The initial AccessMap cost function**

The initial AccessMap cost function was a linear combination of features interpreted as impedences with arbitrary scaling factors and a semi-arbitrary quartic function for evaluating the impact of incline (this cost should be convex).

The exact form of a parameterizable pedestrian cost function is an active topic of research, but AccessMap has explored several forms and can rule out some as non-ideal. Early on, AccessMap implemented a cost function that was a linear combination of pedestrian impedences: incline, crossing the street, a lack of curb ramps, and the length of the path (Figure 10.1). Each impedence had a corresponding arbitrary coefficient that was tuned until routes appeared appropriate to interviewees and the cost function authors. Neglecting the incline impedence, all coefficients were set to 1, so the raw values of edge attributes contributed to the cost function. Incline was first modeled as a quartic function in order to capture the rough avoidance pattern described in user interviews and matching what one would guess about a wheelchair user's experience with inclines: the steeper the incline, the more it should be avoided, i.e. a higher cost should be incurred.

$$\text{cost} = \frac{l}{\text{Tobler}(g, \alpha_1, \alpha_2)} + \begin{cases} \infty & g > \alpha_1 \\ \infty & g < \alpha_2 \\ 0 & \alpha_2 \leq g \leq \alpha_1 \end{cases} + \infty \alpha_3 r + 30c + 60e$$

g is grade (incline))
α_1 is user maximum uphill incline
α_2 is user maximum downhill incline
l is length of the path
α_3 is user requires curbramps
r is indicator of curb ramps
c is indicator of crossing
e is indicator of elevator path

Figure 10.2: **The current AccessMap cost function**

The current AccessMap function is still a linear combination of impedences, but they are now always framed in terms of time (or are infinite in the case of barriers). Scaling factors are real-world estimates for dwell times rather than arbitrary.

The second iteration of the AccessMap cost function made two significant changes: (1) the incline cost was calculated as a piecewise linear convex function centered at -1% grade, with user settings dictating the exact shape of the rate of cost increase, and (2) barriers such as a lack of curb ramps were evaluated to infinite cost for the entire edge weight.

These first two attempts revealed inadequacies in representing pedestrian impedence. First, understanding the relative impacts of each feature and user settings was difficult, as the use of arbitrary scaling constants and incline functions limited understanding the trade-offs. For example, what would it mean to create a quartic function rather than quadratic when comparing incline to sidewalk length? These are not the same "currency" of consideration and cannot be compared directly or intuitively. The second iteration that centered the incline function at -1% addressed a problem raised in informal interviews: most participants preferred a slightly downhill path to a perfectly flat one. While there may be significant variation in preference regarding ideal incline that has yet to be studied, this is clearly a parameter that must be accounted for within a responsible pedestrian cost function.

The current iteration of AccessMap is still a linear combination of pedestrian impedences, but all are framed in terms of time of travel (seconds) and can therefore be compared (Figure 10.2). Hard barriers incur infinite costs and are calculated as early as possible for efficiency, as once a single infinite cost has been discovered, the rest of the function no longer needs to be evaluated. The incline function is an adaptation of Tobler's Hiking Function which, while likely not entirely accurate for real-world pedestrian movement, produces a fairly realistic estimate of movement speeds based on inclines that can then be combined with the length of the path to produce a travel time estimate. Early iterations of the AccessMap cost function added an arbitrary cost for crossing the street to avoid seemingly too-frequent street crossings, but the current version simply adds an estimated time delay to achieve the same (and more realistic) outcome, and could eventually be parameterized by aggregated location data. Barriers are currently set to hard infinite costs. Consequently, current users of AccessMap receive routes that balance speed and accessibility, reflecting the preferences of a wide range of pedestrians.

10.1.2 *Challenges in cost function parameterization: describability*

While it is clear that many environmental features impact pedestrian mobility and accessibility, pedestrian preferences regarding them may be difficult to communicate. Hashemi & Karimi found that automatically-planned accessible pedestrian trips often differ from what is actually preferred by pedestrians with mobility limitations due in part to difficulty in communicating trade-offs and personal preferences[35]. In AccessMap, the cost function is parameterized by maximum uphill and downhill steepness settings, defined as grades (slope, rise over run). However, in informal interviews, pedestrians did not feel confident in identifying these numbers from memory or experience: they estimated steepness visually and by attempting to climb hills. It is unlikely that pedestrians in general can be confident in estimating such a number, but it is essential to have this information in order to parameterize a reasonable cost function.

One potential solution to this problem is to use past experience, examples, and visual feedback, a strategy employed by AccessMap. In order to communicate steepness information, AccessMap colors each sidewalk based on its estimated steepness and the current user's settings:

a relatively flat and easy path is green, a moderate path is yellow, and a difficult path is red (impossible paths are dotted red lines), with intervening inclines interpolated along a perceptual color map. As an AccessMap user changes their settings, the sidewalk colors update. This feedback provides a potential mechanism for dialing in personal settings: someone who has visited any location in Seattle before and noticed too steep of a hill can move the slider until a sidewalk along it becomes a dotted red line. In addition, users could potentially compare Google Street View™ images to AccessMap, comporting a simulated trip with the tool of interest. Eventually, with widespread use, pedestrians may learn their "steepness numbers" as more tools accommodate their needs and experiences.

Another option going forward is to estimate probable default settings for individuals based on questions that act as proxies for cost function parameters. For example, a combination of a self-rating of athleticism, years of using a device, age, or stability may be predictive of uphill and downhill settings. Future work on this subject should include large-scale surveys that compare individualized parameter sets with a wide set of questions in order to identify a minimal set of questions that adequately capture pedestrian mobility profiles.

10.1.3 *A way forward with units of exchange: weighing pedestrian concerns*

Past work has identified that pedestrians with mobility limitations (and likely all pedestrians) weigh a complex set of concerns when deciding on what path is preferable when moving from an origin to destination[35,37]. The primacy of time in the modern AccessMap pedestrian cost function provides a convenient way in which to ask about pedestrian preferences that are otherwise more difficult to communicate and that capture realistic stories and preferences revealed in both informal and formal interviews. For example, when planning a specific trip in the Greenwood neighborhood, several participants indicated bet-hedging strategies: they would attempt a seemingly difficult path but have a backup plan. Similarly, participants would frequently answer, "it depends" to questions about incline, because they would prefer to tackle a steep hill if it saved a significant amount of time. This last point, the idea of trading off time of travel for impedences that are not measured in units of time, provides an opportunity to frame more complex pedestrian cost functions in terms that can be communicated by individual pedestrians: would you be willing to traverse this impedence if it saved five, ten, fifteen minutes?

Thus, we can attempt to incorporate much more complexity and diversity of preferences into future cost functions, creating paths that are not just accessible and fast but also account for:

- Fine-grained speed vs. accessibility preferences.
- Proximity to greenways and a pleasant environment.
- Avoidance of loud or crowded areas.
- A path that includes safety features like lamps at night or proximity to public and populated spaces.
- A path that avoids pollution.

10.2 A PORTABLE TRANSPORTATION NETWORK: REMOVING THE BOILERPLATE WITH ENTWINER

Version 2 of AccessMap was based on an in-memory transportation network model using networkx[31], a research-oriented Python graph library. This approach made flexible routing and initial analyses of urban pedestrian network structure straightforward, but came with limitations in serialization formats and a rapidly approaching explosion in RAM usage. Other routing engines tend to serialize graphs to custom formats[40,46,86], embedding costs directly in an edge list and then reading this minimized edge list into memory, finding a balance between speed of in-memory graph traversal and the size of the resulting graph. In addition, these formats are optimized for project-specific routing strategies and are not amenable to scientific analysis, which would require a deep understanding of the codebase of the associated routing engine. Finally, the need to compute edge costs at runtime for arbitrary inputs necessitates storing network attributes for each edge, significantly increasing the size and complexity of the embedded data. To address these concerns, I created the Entwiner project, which ingests linear GIS data formats to create a portable, embeddable, SQLite[36]-backed transportation network serialization format as well as a networkx compatibility layer for directed graphs.

The use of SQLite provides immediate solutions to the challenges described above. SQLite provides a regular, tabular data format for describing and rapidly accessing data using a common query language (SQL), and is therefore amenable to describing the diverse edge data types required for pedestrian networks. Entwiner stores graph data (edges, nodes) as edge (u, v) and node (n) lists, with covering indexes creating for each variable as well as the (u, v) node pair. Entwiner databases conform to the OGC GeoPackage[14] format, and can be opened, inspected, and edited using standard GIS software. Because the graph data can be immediately serialized and deserialized using any programming language or software tools that can inspect SQLite (and GeoPackage) data, Entwiner databases can be deployed in a wide variety of contexts where a transportation network is needed, including smartphones, smart embedded devices, low-resource cloud computation, and situations where internet connectivity is limited or unavailable.

As a GeoPackage-based network, Entwiner provides three key perspectives for inspecting and deriving spatial network data, all of which were put to use for creating the AccessMap pedestrian network: tabular data (retrieving street geometries from sidewalk rows, curb ramp associations), spatial data (nearest-neighbor queries, Sidewalkify, Crossify), and network (graph) data (Sidewalkify, street intersection identification). Extracting more accurate and topologically correct sidewalks using Sidewalkify required using all three perspectives in concert. Linking these perspectives also assists in addressing common network queries.

Entwiner makes use of these perspectives to provide easily-accessible boilerplate for transportation network analysis and queries. For example, graph analytic questions of spatial networks often start off-graph, requiring a change in algorithmic approach or the creation of a temporary augmented graph that includes new nodes and edges. Shortest path algorithms require in-graph origin and destination nodes, but a shortest path query for requesting directions (A to B routing) is rarely requested from existing nodes in the network; real-world starting and ending locations are not on the embedded graph at all, being neither a node nor an edge. The solution to this challenge shared by virtually all routing engines is to find the nearest network element to the query location (subject to a minimum distance filter) and, if this location is an edge, to implement a routing strategy that effectively embeds a new temporary node at the closest-projected point along that edge as well as two temporary edges connecting the temporary node to the rest of the graph. This strategy combines spatial and network perspectives in order to implement essential and basic functionality for a routing engine, functionality that is typically missing from the core data structures of routing engines and is instead uniquely implemented by each routing engine. With Entwiner, the essential steps of this process become explicit and clear: the nearest-neighbor search is facilitated by a ubiquitous spatial index extension and common distance queries and then either the nearest neighbor algorithm is modified to accept initial and final pseudonodes or an overlay or augmented graph structure is created on which standard node to node shortest path algorithms are run. Entwiner provides an augmented graph, networkx-compatible data structure to assist in the process as well as nearest-neighbor search function that can be easily replicated in any programming language. Similarly, Entwiner offers an efficient and convenient interface for any nearest-neighbor or intersection queries of interest, calculated in real-world units (meters) from geodetic (longitude-latitude) inputs.

The core Python functionality of Entwiner is in its networkx adapter, which implements an SQLite-backed version of a DiGraph class (directed graph), implementing all interfaces. Consequently, the Entwiner graph class is compatible with networkx-based algorithms and can be traversed or analyzed as if it were a "dict of dicts", where asking for a key ("G[1]") produces a dict-like of successors of node 1, and asking for a key of the successor data structure ("G[1][4]") produces a dict-like of edge attributes corresponding to the edge (1, 4). Predecessors can also be accessed from the top-level graph structure, allowing for efficient local traversals. Much of the work presented later is based on these data structures and off-the-shelf graph analysis algorithms, including discussions of graph connectivity and large-scale urban analytics. Entwiner additionally serves as the core data structure and interface for the routing engine used by AccessMap, Unweaver.

While Entwiner addresses many challenges and has practical use deploying transportation networks and research, it has known limitations and future work that would significantly improve its functionality has been identified. One current limitation is that it produces a unique edge list (u-v pairs) table, with a strict "UNIQUE" SQL constraint, which is only compatible with graphs or directed graphs, but not multigraphs or multidigraphs. This is a potentially significant limitation for more complex networks, including more detailed pedestrian networks, as they may actually be multigraphs or multidigraphs. For example, a footpath may temporarily split into unidirectional pieces, such as when a joint footpath and cycleway are split, in which case two edges should share a u-v node pair. Entwiner should therefore be updated in the future to optionally remove the uniqueness constraint and to provide a networkx-compatible MultiDiGraph interface. There is also an opportunity to promote Entwiner or a solution derived from it as a routing extension to the OGC GeoPackage standard. Because the modifications to tables are relatively simple and straightforward, primarily implying the addition of "_u" and "_v" node columns to geometry tables and a separate nodes table with an "_n" column of unique node IDs, this extension would be language-agnostic and could serve as a common format for building, sharing, and analyzing transportation network data.

10.3 FLEXIBLE ROUTING WITH UNWEAVER: A RESEARCH-ORIENTED ROUTING ENGINE

10.3.1 *Limitations of existing routing engines for pedestrian directions*

Reviewing open source routing engines for use by AccessMap revealed severe limitations. Most open source routing engines focus on OpenStreetMap data and are designed with street traffic in mind, resulting in implementations that were brittle to modifications relevant to pedestrian needs. For example, OpenTripPlanner[46] is focused primarily on transit routing on streets, heavily modifying the underlying network during the process of importing OpenStreetMap data and distributing the task of storing and maintaining edge attributes among many different modules. In addition, providing the user-parameterized settings required to currently capture cost function preferences of diverse pedestrians would have required the definition of server-side configuration options as well as web API arguments on a per-attribute basis. While it is possible to represent the pedestrian network within OpenTripPlanner and there are ongoing efforts to incorporate more inclusive pedestrian experiences within the project, these limitations make it difficult to use for prototyping, research, or highly-flexible dynamic cost functions.

Other routing engines demonstrated similar limitations. Some routing engines exclusively offered static edge weighting[86], making the user-parameterized cost functions required for pedestrian routing impossible. All stored their graph representations in bespoke and uninterrogable serialization formats, frequently focusing on optimizations for more efficient shortest-path finding on streets. Consequently, existing routing engines would have required a deep investment writing new functionality and modifying the inner workings of a large and dense software package in order to implement the basic functionality required for AccessMap or even most graph analytic algorithms.

Early implementations of routing in AccessMap used the networkx library, a research-oriented network library for Python, as it implemented flexible data structures and graph analytic algorithms of interest to the project. Many of the challenges of using networkx for this purpose were addressed by using the Entwiner library, but several remained. AccessMap needed to implement several cost functions, with some pre-parameterized for static weighting (stereotyped profiles for analysis purposes) and others using cost functions parameterized by users at runtime.

These cost functions would potentially need to be written and parameterized by researchers with familiarity with Python, but not necessarily experts in software engineering. Multiple profiles may share cost functions, but simply provide different static weights for certain inputs. Each profile would need to describe the cost function parameters available to users, imply the construction of a web API, and validate user inputs to produce meaningful error modes for client applications. Finally, one of the limitations of alternative open source routing engines was the inflexibility in results of web API requests, such as those for directions, as they were all intended for turn-by-turn street navigation rather than pedestrian navigation, both including unnecessary of misleading information and neglecting necessary data, such as edge attributes.

Unweaver is a routing engine focused on flexibility. It can read many data formats (including that of OpenStreetMap), find shortest-path routes via a web API, and allows completely customizable combinations of cost functions and directions specifications, summarized in JSON "profiles". Unweaver's costing strategy includes dynamic (as opposed to precalculated, static) edge costs for when profiles need to be heavily parameterized on a per-user basis.

10.3.2 *Profiles*

In Unweaver, a routing profile is a combination of a cost function and a directions generator. The cost function defines how much of a penalty is incurred when traveling along a particular edge and the directions generator creates JSON that describes the results for API consumers, e.g. turn-by-turn directions.

An Unweaver profile is a JSON file that references:

1. A profile name.
2. (optional) A cost function: a (relative) path to a Python script defining a function that accepts any user-parameterized arguments and returns a function that accepts edge data and returns a number (the cost function).
3. (optional) A direction functions: a (relative) path to a Python script that accepts data from the shortest-path directions endpoints and produces a user-defined JSON-compatible output.
4. (optional) an arguments ("args") object that defines typed web API arguments that can be passed to the cost function.

Only the profile name is required, with all other profile sections having reasonable built-in defaults. The cost function defaults to returning edge length that is calculated during graph creation, returning a literal shortest-path cost function. The directions function defaults to a ordered edges (including edge properties) and the overall path geometry. By default, there are no user-defined arguments.

10.3.3 *Customized costing of shortest paths*

As discussed earlier, one of the chief limitations of most routing engines is that edge costs are baked into the graph: all costs must be precomputed, limiting flexibility during runtime. While a baked-in approach can be efficient for calculations (and Unweaver does allow precalculated costs), there are situations where it is inadequate for tackling a particular shortest-path challenge. For example, the length of time it takes for a car to traverse a street is usually embedded as a number within the routing graph, not calculated at the time of finding the shortest path. Within the AccessMap project, the need for runtime costing is clear: pedestrians have a wide range of preferences that are not easily summed up in one or two preset costs, and requires personalizing cost functions on a per-user basis.

Unweaver achieves flexible, dynamic costing by leveraging Python: all cost functions are defined as simple Python functions that can be instantiated at runtime. This means that a user of Unweaver can include any of the wide range of numerical and scientific libraries available in the Python language within user-defined cost functions that are parameterized through web APIs at runtime: costs are calculated on the fly using a single Python function.

Unweaver cost functions are directly compatible with networkx cost functions: they take in three direction parameters: start node (`u`), end node (`v`), and dict-like attribute data (`d`) and return a cost (number) representing the penalty of traversal (return `None` for an infinite cost / non-traversable edge).

10.3.4 *Directions*

There is no standard for the format in which directions requests should be formatted, so Unweaver also allows this to be user-customizable via a directions function. A directions function, like the user-defined cost functions, is a single Python function. A directions function is specified to receive a list of node lists (i.e. one or more paths) along with a list of edge lists (the edges traversed by each path) and return a JSON-compatible Python object. Unweaver provides a default directions function that produces a JSON result similar to the Mapbox directions API.

10.3.5 *Web API*

For each profile, Unweaver automatically generates web APIs for requesting point-to-point directions (A to B routing), a shortest-path tree, and “reachability” from a point source. Cost function parameters sent via the web API are automatically validated using the Marshmallow framework, which enforces (flexible) typing using Unweaver profile definitions. Once validated, these data are used to parameterize the user-defined cost function generator referred to in an Unweaver profile, creating a custom cost function for use during the request. This cost function is then made use of by the aforementioned graph traversal algorithms to return (customizable) directions results, shortest-path tree results, or reachable space results in JSON format.

CHAPTER 11. THE PROMISE OF USING THE PEDESTRIAN NETWORK TO UNDERSTAND AND IMPROVE THE BUILT ENVIRONMENT

Several different metrics of pedestrian accessibility have been proposed, including density of amenities like sidewalks and crosswalks[16], density of conditional barriers[45], and street network analyses[21]. However, these metrics do not directly model pedestrians' experiences: they do not model the step-by-step path that pedestrian can actually take, along with the conditional barriers that may be met along the way. This results in the systematic undercounting of pedestrian accessibility. For example, two neighborhoods may have the same density of sidewalks, but may differ in their connectivity, meaning they should receive different accessibility scores. Similarly, the density of raised curbs or curb ramps may be identical between two areas, but due to their distribution on the actual paths, one neighborhood may be fully divided along a street lacking curb ramps, whereas another is not.

Modeling the pedestrian experience requires accounting for the aspects of the pedestrian environment that determine whether pedestrians can and will use a space for pedestrian travel: on foot or using assistive devices. Whether a space can be used by a given pedestrian depends on the (mis)match between that individual's capabilities and preferences and the environment itself, which may act as a barrier to travel for many. Importantly, these spaces must be connected: there must be a path from a pedestrian's start point to their destination free from barriers that prevent that individual's access. For example, while sidewalks are often considered an improvement to pedestrian access, if they are too steep, they are part of a barrier to access for many wheelchair users. Even if there is only one sidewalk that is too steep along a particular path, the entire path should be considered inaccessible: it cannot be used to reach that destination for that pedestrian. Therefore, modeling pedestrian access requires modeling connections, for which graph models are suitable.

A pedestrian graph network is analogous to a transportation network used for automotive traffic: edges are potentially-traversable path segments and nodes are "intersections" where pedestrian paths meet. For automotive transportation networks, there are very few street attributes that conditionally prevent access, one example being a maximum height limitation that impacts large vehicles. In contrast, frequently-encountered attributes of a pedestrian network can entirely prevent access to many populations. In addition, path types associated with safety and pedestrian preference can be entirely missing, such as sidewalks. Modeling pedestrian access using a network approach therefore presents a unique, multi-dimensional challenge: what pedestrians are being modeled (normative, stereotypical wheelchair user, etc), what are the possible paths that pedestrian could take to or from certain locations, and what would qualify as "good" access?

11.1 CURRENT APPROACHES TO UNDERSTANDING PEDESTRIAN ACCESS

11.1.1 *Manual curation and surveys*

A common view of sidewalk availability or pedestrian access is asset-centric, i.e., it counts the existence of assets like sidewalks or crosswalks or points of interest considered relevant to pedestrians such as grocery stores. Some approaches, like the popular WalkScore[9], do not incorporate sidewalk presence into their calculations whatsoever, and rank "walkability" using street data and amenity information, resulting in an image (raster) that scores multi-block regions of an area as being relatively walkable (or not). Other attempts to assess sidewalk availability simply count the number of sidewalks in a region of interest, such as a neighborhood, sometimes normalizing by the number of streets, resulting in a percentage of streets that have sidewalks on at least one side. Finally, some approaches attempt to measure sidewalk density as a metric of pedestrian access[58,87], converting sidewalk lines into a heatmap where anywhere within 10-20 meters of a sidewalk is considered "served by sidewalks" and any area that's significantly farther away is not.

Asset-centric measures of pedestrian accessibility, or even just sidewalk access, give a misleading impression of pedestrian access. In an asset-centric view where the area is considered accessible because 99 percent of streets have sidewalks, there may be no curb ramps, or the sidewalks may be too narrow at an important corridor connecting commercial zones, or there may be a dividing hill that breaks the space up into three usable pedestrian spaces. Whereas these challenges will be missed by an asset-centric view, they will be captured in a network-based analysis, which evaluates every asset as it is traversed.

11.1.2 *Connected network analysis*

Figure 11.1: **Connectivity analysis in central Seattle**

Connected subgraphs within the Seattle pedestrian network subject to a single constraint: the presence of curb ramps. Each color represents a different subgraph that is disconnected from all other subgraphs.

A simple graph analytic question that can be asked of a pedestrian network is connectivity. A graph is considered connected if all of its nodes can be reached from any other node via its edges. By inducing realistic pedestrian constraints on the pedestrian network, connected subgraphs (disconnected from one another) can be elucidated, such as in Figure 11.1, where numerous subgraphs appear once a curb ramp constraint is added. Under this constraint, Seattle is made up of three large connected graphs (the Magnolia neighborhood, West Seattle, and most of the remainder), but a large portion of isolated subgraphs are generated, disconnected from the rest of the network. A high proportion of these isolated subgraphs, which can be thought of as pedestrian islands for pedestrians that require curb ramps, could be a useful indicator of accessibility and network disconnections.

11.1.3 *Walksheds*

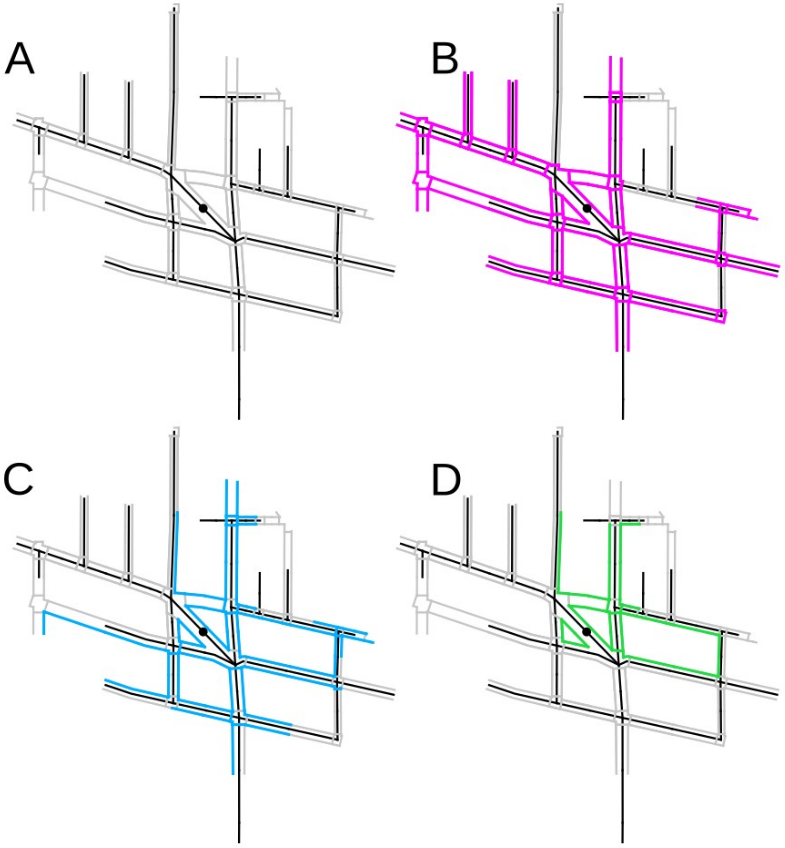


Figure 11.2: **Walksheds in the Fremont neighborhood of Seattle**

Walksheds analyze the network elements reachable subject to a constraint, commonly a quarter mile (roughly 400 meters). For all subfigures, black lines represent a walkshed on the street network subject only to a 400 meter maximum distance constraint. (A) An unconstrained pedestrian cost function generates the light gray extent shown. (B) A cost function limited by inclines greater than 8.3% generated the fuchsia walkshed. (C) A cost function limited by a requirement for curb ramps generated the blue walkshed. (D) A cost function limited by a requirement for marked crossings generated the green walkshed.

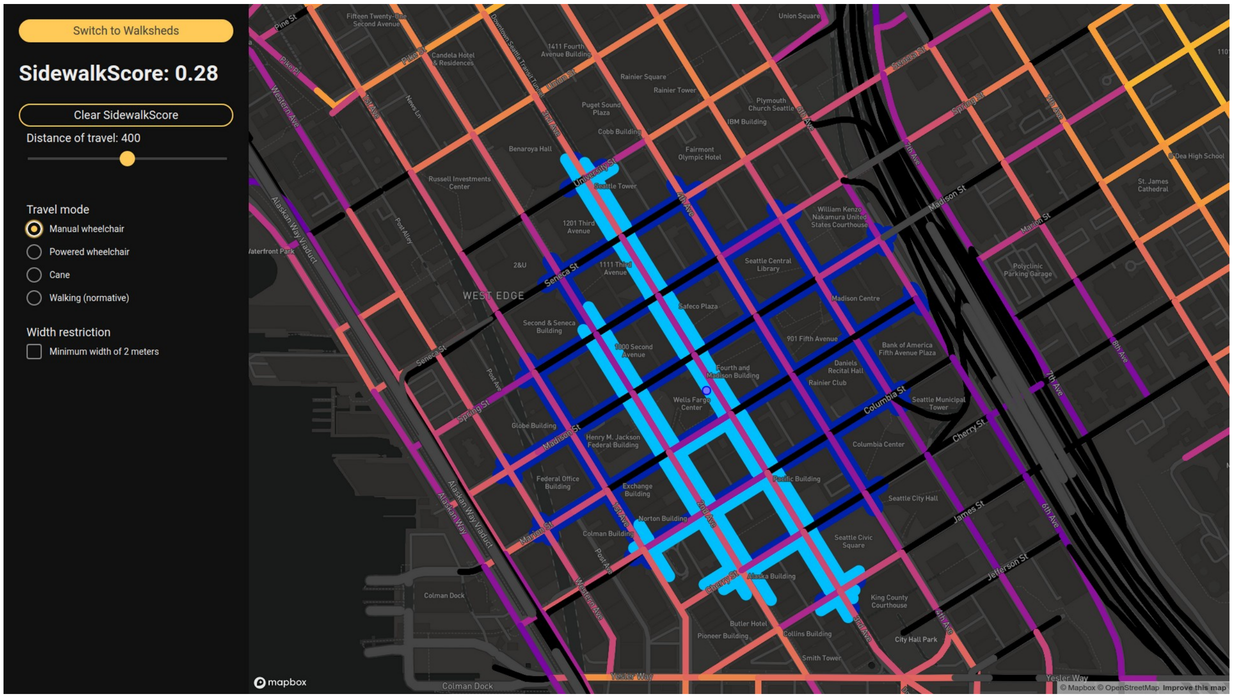
Walksheds are a common, network-based analytical tool in city and transit planning that estimates the locations that can be reached given a starting location and maximum distance or time. For example, transit planners in the United States often discuss a quarter-mile walkshed service area for a particular transit stop, indicating that persons within that walkshed can get to and from that bus stop by walking a quarter of a mile. However, the accuracy and utility of a given walkshed is predicated on the method by which reachability is considered. Some walksheds model service areas “as the crow flies”: they draw a circle, often a quarter mile in radius, out from the transit stop, which leads to absurd implications, such as a bus stop servicing the middle of a lake or pedestrians that can walk through walls. A more sophisticated approach to walksheds models pedestrians using the street network, wherein pedestrians are modeled as if they were slowly-moving cars. The resulting set of reachable street lines that looks more accurate, but is still not describing pedestrian mobility: putting aside the mobility profiles of individuals with limited mobility, most pedestrians move more slowly on inclines and will avoid busy streets if there are no sidewalks.

Figure 11.2 shows walksheds induced on the pedestrian network under various pedestrian cost function profiles, with street network walksheds (black) for comparison. A key advantage of walkshed analysis is the questions of reachability enabled by the network and shortest-path analysis. For example, subfigure B in Figure 11.2 shows that several sidewalks and crossings are unreachable under an incline constraint that matches ADA guidelines for ramps. Many of the elements shown as unreachable are, individually, much flatter than this constraint, but cannot be reached within a 400 meter path without traversing a sidewalk that is too steep.

11.2 METRICS OF TRAVERSABILITY AND ACCESS: SIDEWALKSCORE, EQUITYSCORE

11.2.1 *SidewalkScore*

SidewalkScore is a score of sidewalk accessibility that compares reachable areas, or walksheds, between the pedestrian network and the street network, conditioned on a particular mobility profile (Figure 11.3). SidewalkScore leverages a real pedestrian transportation network, the same one used by AccessMap, to create walksheds that are conditioned on stereotyped mobility profiles. The sum distance of reachable paths is calculated for this walkshed and represents the absolute pedestrian access of that point in space according to a particular mobility profile. If AccessMap’s pedestrian network is dense and accessible at that point of interest, the walkshed will be roughly circular or ball-shaped and have a large sum of path lengths, and if the pedestrian network is not accessible, the walkshed might have a non-circular shape. For example, a walkshed for a manual wheelchair user in downtown Seattle is typically elongated in the Northwest/Southeast directions and limited in the Northeast and Southwest directions due to extreme inclines. However, because our pedestrian network is only aware of sidewalks and street crossings, this scoring can not be used alone to gauge sidewalk accessibility: a non-dense areas with few streets but good sidewalk coverage should receive a good score, but would receive a low walkshed distance. Therefore, SidewalkScore normalizes the pedestrian walkshed by a street walkshed, wherein the same mobility profile is used to create an analogous walkshed that used only streets, ensuring that the availability and properties of sidewalks are the primary difference between any point sampled near a street.



SidewalkScore: 0.28

Distance of travel: 400

Travel mode
- ● Manual wheelchair
- ○ Powered wheelchair
- ○ Cane
- ○ Walking (normative)

Width restriction
- ☐ Minimum width of 2 meters

Figure 11.3: **SidewalkScore dashboard**

SidewalkScore dashboard showing the two walksheds used to calculate a given SidewalkScore, one on the pedestrian network subject to a pedestrian cost function, the other an unconstrained walkshed on the street network. The dashboard allows for an intuitive sense of the score: the areas that appear reachable by a pedestrian take up approximately the fraction of space relative to the street network walkshed as indicated by the SidewalkScore (in this case 0.28).

SidewalkScore can be evaluated at any point in space so long as corresponding streets and sidewalks can be located. In Figure 11.4, a city-scale evaluation of SidewalkScore for every street in the Seattle Street Network Database dataset is displayed on top of WalkScore, demonstrating disagreement between pedestrian network access and metrics based on amenity density.

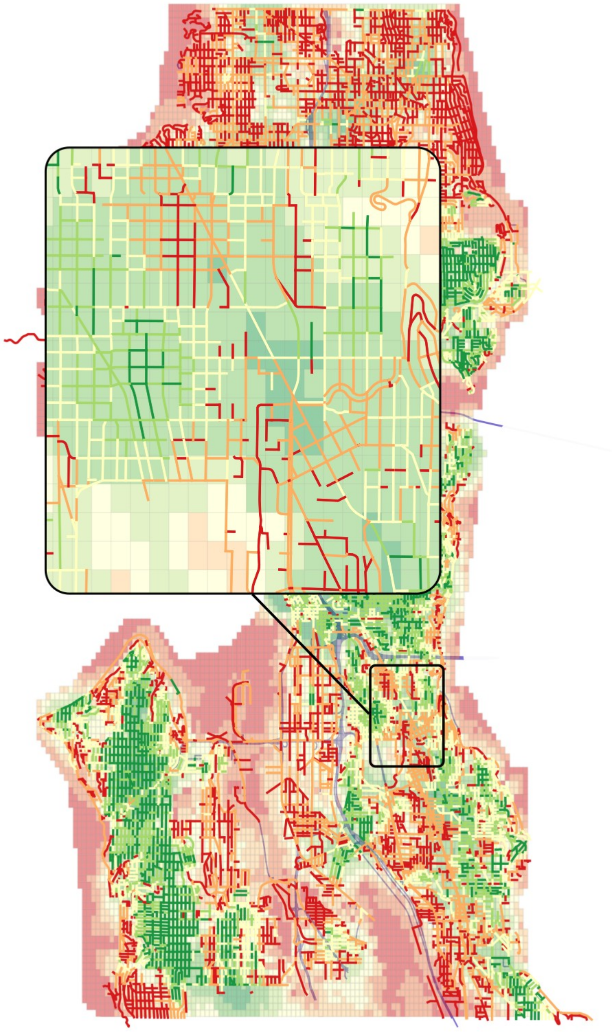


Figure 11.4: **SidewalkScore in Seattle vs. WalkScore**

An overview of SidewalkScores from a normative walking profile versus WalkScore cells. WalkScore frequently gives good scores to areas that lack sidewalks and have hills that are too steep for many pedestrians.

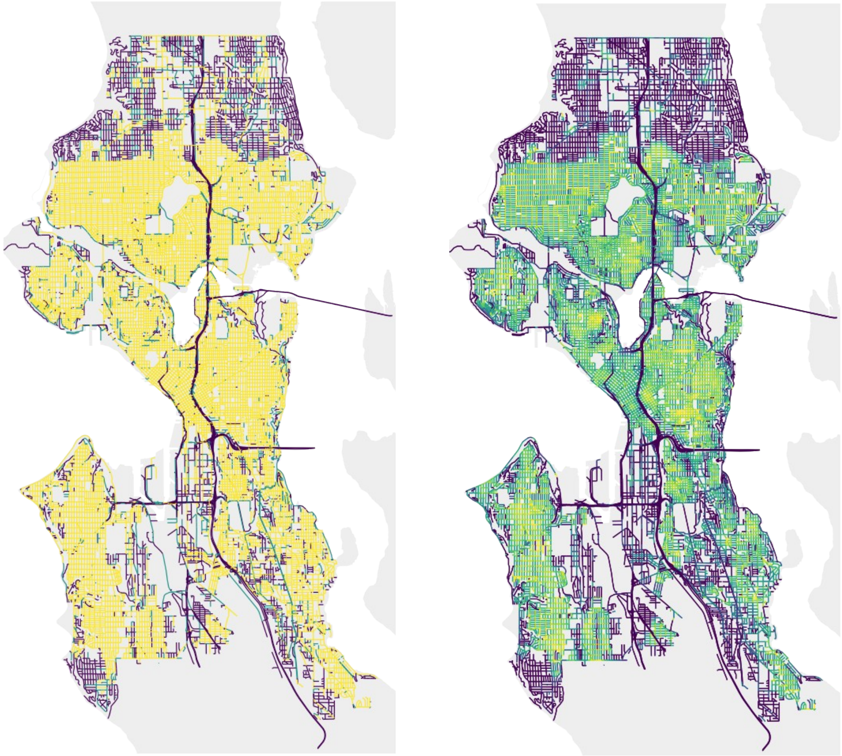


Figure 11.5: **Sidewalk count vs. SidewalkScore**

On the left is a visualization of the number of sidewalks per street in Seattle, WA. Purple streets have no sidewalks, turquoise have one, and yellow have two. On the right is a visualization of SidewalkScores for every street in Seattle under an unconstrained pedestrian profile. SidewalkScores are on the same color map, with 0 being purple and 1 being yellow.

Agencies frequently use sidewalk count and density as a metric for pedestrian access, which lacks network information for evaluating the impacts of infrastructure availability on actually maneuvering the space as a pedestrian. Figure 11.5 shows a comparison between sidewalk density and SidewalkScore for Seattle. The SidewalkScore shown is under a relatively unconstrained profile with incline settings greater than 15% grade and no barrier constraints. Because simple counts and density overlook network information, they can easily overestimate the accessibility of a pedestrian space regardless of individual pedestrian needs or preferences.

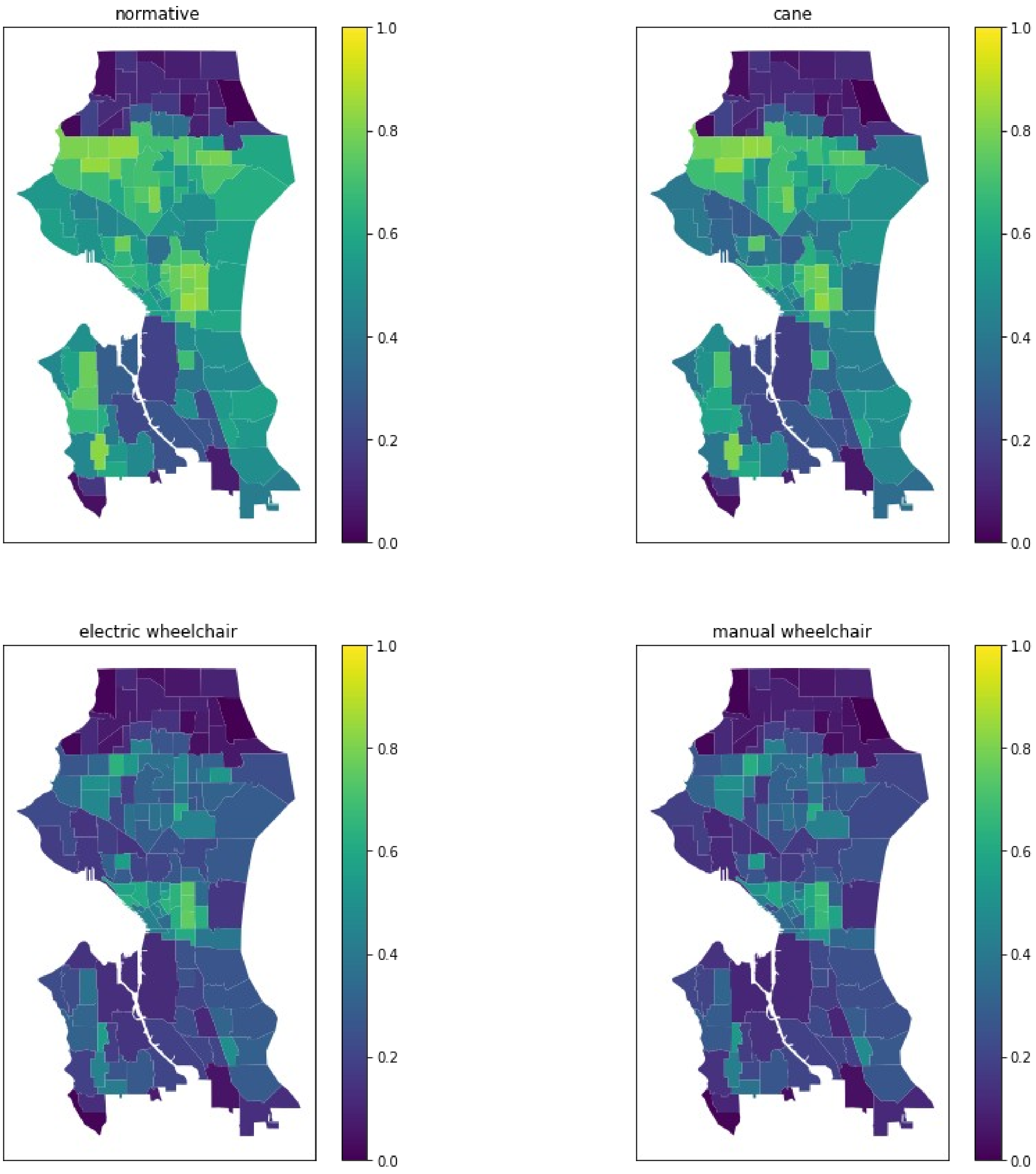


Figure 11.6: **SidewalkScore for different profiles aggregated by census tract**

SidewalkScores are contingent on pedestrian preference profiles. Each panel title indicates the profile used for evaluation, with stereotyped manual wheelchair, powered wheelchair, and cane profiles compared to a "normative" unconstrained profile.

Because SidewalkScore is contingent on the pedestrian profile used for evaluation, it can be used to analyze a city’s accessibility on an individual basis. Figure 11.6 shows SidewalkScores for four stereotyped profiles, aggregated by census tract (means are shown). Each profile generates a unique SidewalkScore visualization, showing disagreement between stereotyped pedestrian groups on the accessibility of various neighborhoods. As better and more accurate pedestrian clusters are identified (building on work in Chapter 4), SidewalkScore will be able to evaluate accessibility to known quantities of pedestrian populations, providing insights for policymakers, bureaucrats, and advocacy groups.

11.2.2 *EquityScore*

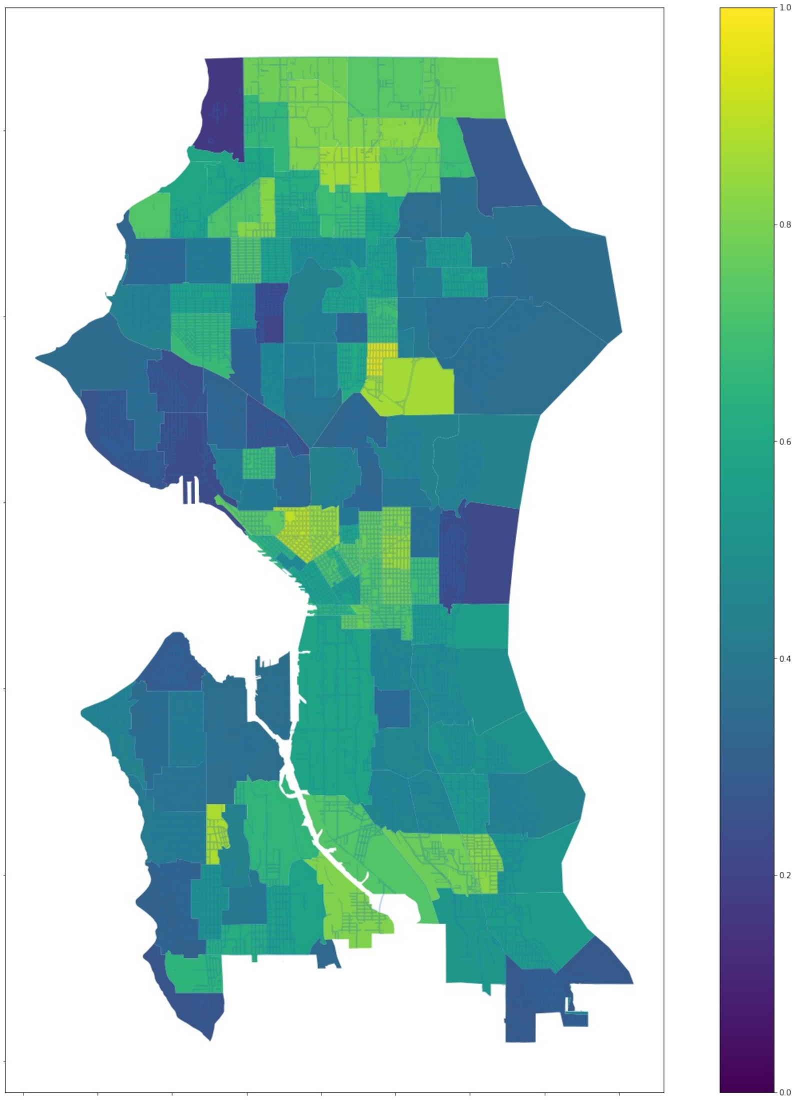


Figure 11.7: **EquityScore per the stereotyped manual wheelchair profile**

EquityScores by census tract in Seattle. An EquityScore of 1 indicates parity between profiles whereas a score of 0 indicates that wheelchair users have a much more limited range within a given region.

The SidewalkScore approach generates a normalized numerical evaluation of the space that can be reached by a particular pedestrian preference profile, giving a sense of the overall traversability of a given space, street, or area around a point of interest. This information can be used to understand where pedestrian accessibility is overall poor for a single stereotyped pedestrian, but does not state whether this accessibility implies inequity between different “types” of pedestrians. EquityScore attempts to address this problem by comparing the SidewalkScores of two profiles, namely a stereotyped profile of interest and an unconstrained pedestrian profile. By dividing the constrained profile SidewalkScore by an unconstrained SidewalkScore, the relative accessibility can be evaluated. Figure 11.7 shows such an EquityScore calculated for a stereotyped manual wheelchair user, demonstrating that some census tracts have near parity (University District, near the upper middle, yellow) whereas others show extreme inequity, such as parts of the Greenwood neighborhood (to the Northwest of the University District, dark blue). Through use of analysis techniques like EquityScore, which can be evaluated on the basis of any point in space, equity questions can be investigated regarding access to transit stops, commercial areas, and entire residential neighborhoods.

11.3 ALTERNATIVE GRAPH ANALYTIC INTERPRETATIONS

11.3.1 *Isochrones and advocacy*

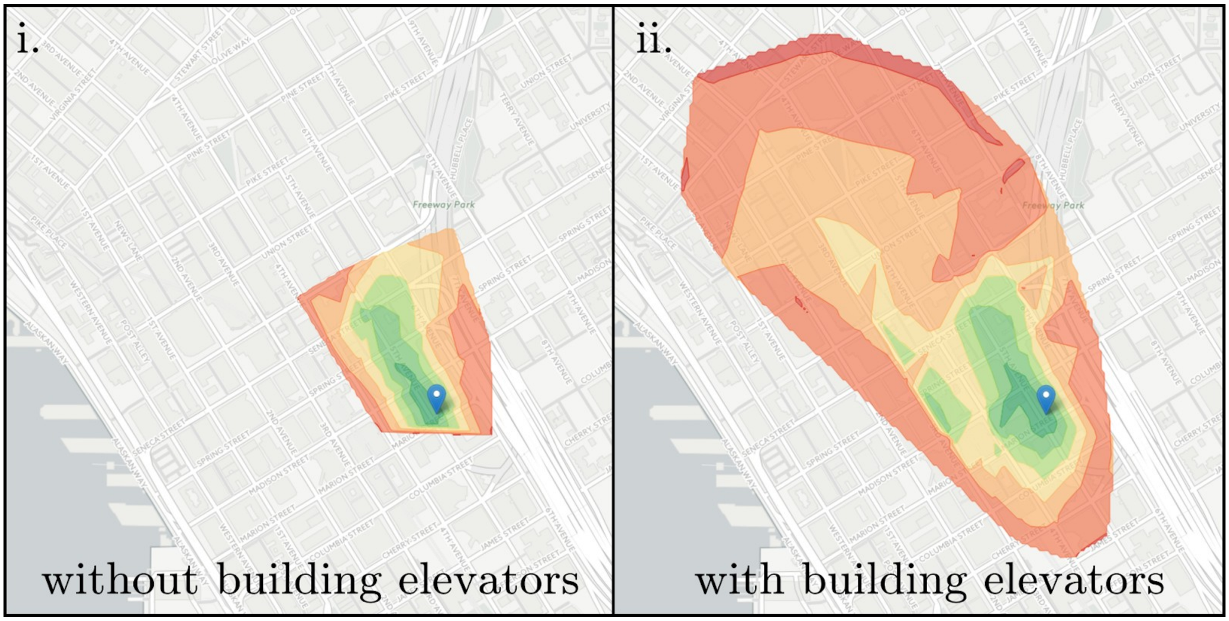


Figure 11.8: **Isochrones in downtown Seattle used for policy advocacy**

(i) An isochrone map derived from an early version of the AccessMap routing engine. Extent displayed reflects the use of a stereotyped manual wheelchair profile without access to building elevators. (ii) An isochrone at the same location and with the same input data, but with allowed use of elevators within public buildings. A much wider area is reached and many spaces become easier to reach than intervening ones, creating “islands” of accessibility by traversing buildings.

Isochrones represent a form of analysis that is similar to walksheds, but retains data on the “cost” of reaching a particular area of interest. Isochrones are derivations of a shortest-path tree where the lowest cost of travel to nodes within the network are retained during traversal using a Dijkstra-like approach. These points are then interpolated into polygonal (or similar) data structures using contouring algorithms. Points sharing the same contour level should have roughly the same difficulty of being reached from the starting point of a given isochrone, so by differentially coloring isochrone polygons, the overall reachability of contiguous network spaces can be estimated. Because isochrones are based on the same network traversal approach as routing, they can be calculated under differently-constrained pedestrian cost functions. Figure 11.8 shows two isochrones calculated for downtown Seattle, starting near city hall, that were conditioned on a stereotyped wheelchair profile. The left isochrone shows that a relatively limited area can be reached under typical conditions due in large part to the extreme inclines to the Southwest and Northeast of the starting location. The right isochrone shows the pedestrian spaces reachable when public buildings with elevators are added to the model, simulating that the buildings are open at a particular time of day. A much wider area can be reached, with shortcuts through buildings creating new accessible “islands” several blocks away. The newly-reachable spaces are a major transit hub and these figures were used to advocate for allowing individuals with mobility impairments to travel through public buildings long after those buildings would otherwise be closed.

CHAPTER 12. LOOKING FORWARD: COORDINATING PEDESTRIAN NETWORK DATA SHARING BETWEEN DIVERSE STAKEHOLDERS

12.1 DATA INTEGRATION IN THE SOCIAL REALM: ACCESSMAP

Using the case study of pedestrian-focused mapping and routing, this work illustrates the need for a data commons approach to data integration solutions, specifically within the subspace of data problems benefiting from public feedback that inform and benefit the social good.

Progress in integrating pedestrian geodata has been hindered by the lack of technical and organizational infrastructure to facilitate cost-effective data sharing between stakeholders. In many instances, institutional barriers stymie data sharing and standardization even though all entities stand to gain by sharing; in the United States, sharing information across organizational boundaries has been a central objective in efforts to improve government operational and service efficiency [38].

Data integration, or the ability to effectively share data among entities and applications and integrate the data efficiently and flexibly [19,32], poses another key problem in private and public service innovation. Namely, stakeholders must currently maintain independent layers of data in addition to a shared data commons. With regard to pedestrian information, this may be due, in part, to the numerous independent stakeholders that maintain and consume sidewalk data and the technical challenge of requiring much more detailed and accurate data than the street network.

Municipal agencies maintain pedestrian data for city planning and compliance purposes, for example, compliance with the Americans with Disabilities Act (ADA)[71]. Regional and national agencies maintain and consume pedestrian data to offer accessible public transportation options and to enforce accessibility legislation. The public consumes sidewalk data for wayfinding applications and to report local problems, such as deteriorating infrastructure or unpublicized barriers, and engage with their communities. Corporations provide analytics and services based on pedestrian activities, such as walkability scores.

However, each of these stakeholders has specific data custody requirements. For example, government agencies and companies providing services based on these data must be able to provide guarantees of accuracy to address liability concerns [43]. Despite distinct data custody and provenance requirements, an accurate, standardized, open, and shared pedestrian geodata layer can greatly benefit all stakeholders, and we envision emerging technologies as being able to engage the broad community to satisfy the needs of validation, data custody and data integration.

12.2 A DATA COMMONS MODEL FOR MULTI-STAKEHOLDER PEDESTRIAN DATA

In recent years, there has been growing interest in the creation of digital data commons, where datasets and supporting software platforms are shared among varied stakeholders. Data commons vary in function and domain, ranging from scientific repositories for research[15] to civic directories for the provision of social services[6] to data pools for predictive business analytics[64]. Regardless of this diversity, organizations involved in the formation of digital data commons face shared technical and social challenges. Technical challenges include interoperability, storage, latency, security, and reliability.

With regard to social challenges, early discussions of natural resource commons focused predominantly on the “free-rider” problem and “the tragedy of the commons”: the idea that people are disincentivized to invest in the maintenance of resources that can be freely obtained and will therefore squander those resources. Although much work has largely displaced notions of the tragedy of the commons with nuanced studies of the multiple modes of investment in common resources that can and do actually exist in the world[24,50,55], challenges certainly remain. For example, Frischmann et al.[24] articulate a number of “social dilemmas” that accompany the creation of knowledge commons, of which data commons can be considered a subset. These include the: production of resources to be shared in the commons; coordination of resource contributions from various stakeholders; definition of the community’s boundaries; and navigation of the entanglement of non-competing data resources with competing resources like time, labor, and hardware.

Many of these challenges - such as the definition of community boundaries and the navigation of competing resources - are specific to the constellation of actors involved in any given data commons and have local solutions that are uniquely contingent upon their patterns of interaction. In other words, “the institutionalizing forms [of data commons] are highly dependent on culture and on people’s interests”[44]. While recognizing the nuances specific to each unique data commons, our work developing AccessMap has led us to observe several design elements that we believe can serve as a model to partially address the aforementioned challenges in a nascent data commons. These elements particularly apply to data commons that are both in the public domain and whose validity can be vetted by an untrained public, while allowing stakeholders some degree of control over their own datasets.

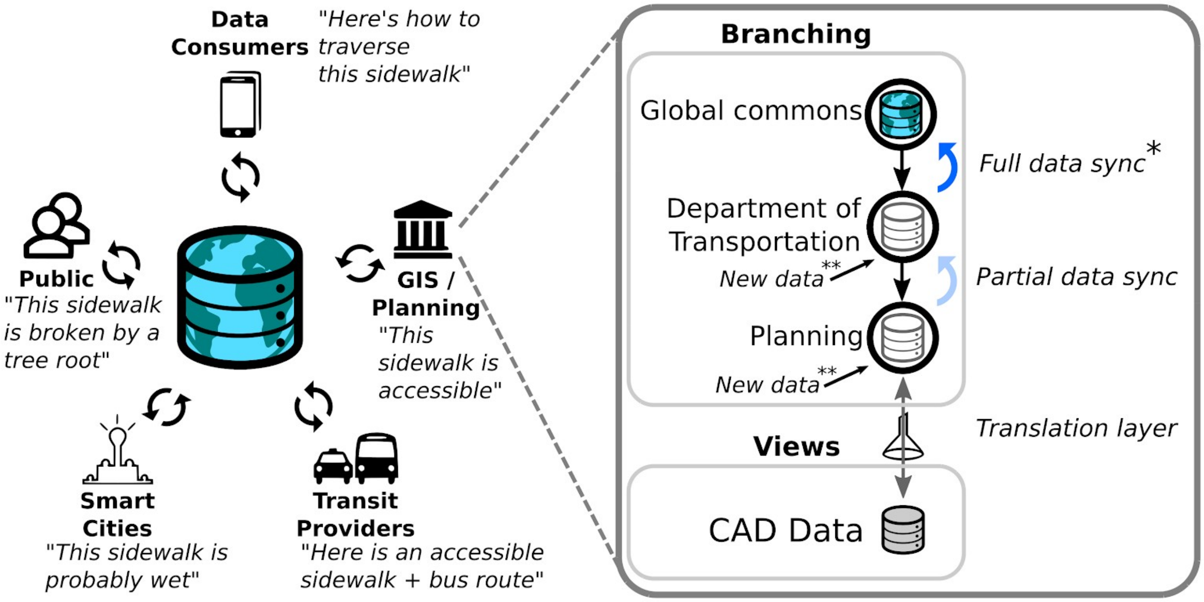


Figure 12.1: **The community-mediated data commons model for data integration addresses many challenges inherent in open sharing, maintaining, and governing data integration among multiple stakeholders with heterogeneous data.**

Despite a constellation of stakeholders with diverging interests, synchronizing with the data commons enforces a standard for data exchange and acts as a medium for community mediation. The inset on the right shows an example of how a government institution can interface with a commons dataset: through large-scale copying (branching), through synchronizing (potentially filtering private data), and through views into existing resources (such as CAD (computer-aided design) files for transportation planning). The * symbol indicates a step where community vetting occurs, and the ** indicates where community-supported tooling can impose structure on new data.

Our preliminary model (Figure 12.1) includes the following components that will be further articulated: (1) a stable data commons to facilitate interoperability, (2) processes for community-governed verification of contributed data (including validation tools for an untrained public to vet the data), and (3) a capacity for customization and reservation of private data layers from the commons. We discuss how the intersecting parts of this model are coalescing within the OpenStreetMap community, and how our experience using OSM to build AccessMap contributes to an understanding of the challenges and opportunities involved with developing this community-supported data commons model. Examples of additional situations that may fit these criteria are nautical maps, building usage, and emergency resource coordination, as well as non-geospatial data, such as government forms and document translation.

12.3 A DATA COMMONS BASE LAYER

OpenStreetMap has developed a robust community of individuals, private companies, NGOs, and agencies that contribute data and tools to a sustainable, current global estimate of the geographical state of the world. Essential to the success of OpenStreetMap is the reliance on an extensible, layer-less data model used in a versioned, global record of the current state of the OpenStreetMap database: if an individual notices something is incorrect, it can be edited and almost immediately changed using community-contributed validation tools.

12.4 STAKEHOLDER VERIFICATION OF CONTRIBUTED DATA

Despite the apparent risk in trusting anonymously submitted map data, stakeholders in the OpenStreetMap community have developed practical verification frameworks that address this issue. A common solution is to create a local copy (or branch) of the OpenStreetMap database and systematically screen all incoming changes. Companies offering driving directions employ this approach, as does the public transportation trip planner offered by the TriMet transportation agency, which is responsible for public transportation in the greater metropolitan area of Portland, OR. This model of periodic review and synchronization promotes contribution to the OpenStreetMap data commons while allowing stakeholders to offer the assurance of accuracy and verification. The right side of Figure 12.1 illustrates a use case for government institutions to keep a locally vetted copy that contains only data validated within their organization.

The Portable OpenStreetMap (POSM) project developed for the American Red Cross takes the concept of a local copy a step further by spawning large portions of the OpenStreetMap software ecosystem as local instances. Intended to gather data in low resource areas where electricity and internet connectivity is scarce, the POSM model for data contribution features delayed synchronization: a subset of OpenStreetMap data is downloaded to the local instance, numerous methods for inputting new data to the server are employed, and the new data can be uploaded to the global OpenStreetMap. Due to the time delay, conflicts can appear and must be resolved. This model resembles branching in version control of software, and the developers of POSM have posited a git-like comparison method for use in their synchronizations[23].

12.5 CUSTOMIZATION AND PRIVATE LAYERS

Essential to many stakeholders is the ability to reserve a subset of data from synchronization. For example, a transportation authority may wish to keep survey data about individuals private for ethical and liability reasons. Or, a stakeholder may want to store information that is inappropriate for OpenStreetMap, such as data on temporary disaster conditions, but still needs to synchronize other information with the data commons. Therefore, these stakeholders require a systematic method to prevent synchronizing potentially arbitrary categories of information. While many potential methods can achieve this goal (e.g., unique tags for reserved data, a tracking table for private data), there is currently no readily available system for selectively synchronizing OpenStreetMap data. In addition, many of the data producers do not have the resources to perform stringent data processing, which, according to some estimates, consumes 70% of the time spent on analytic projects[13].

12.6 CHALLENGES AND OPPORTUNITIES IN THE EMERGING DATA COMMONS MODEL

Our experience with AccessMap makes clear that mutually beneficial integration of pedestrian data among many stakeholders has challenges, some that are shared across all data integration problems and others that are unique. We propose that the problem of the pedestrian information gap is just one instance of a class of data challenges relating to social good that dictate the need for a community-mediated data commons approach (Figure 12.1).

Work in data federation and data integration highlight some of the common challenges that enterprise data integration efforts experience, particularly as they attempt to extract enterprise-specific value from analyzing massively heterogeneous available data[24,30,32]. A recently proposed model calls for a centralized “data lake” repository that contains vast amounts of raw data kept in its original format and dataset affiliation and that is available on demand[13]. But “raw format” data are rarely ready for immediate consumption without processing; therefore, a repository that does not address data staging only defers the problems of dealing with data quality, provenance and governance to the entity accessing the data from the data lake. In an enterprise solution setting, Terrizzano et. al. propose that a team of curators could provide up-front services in both data cleaning and arbiter licensing and permissions guidelines. In the scenario we described, the central repository (the data commons) has no employees or resources to task a group to vet data contributions[60].

If stakeholders could easily maintain private copies of data, we risk creating an ecosystem of data silos with no incentive for sharing information. This risk is an open question for the plausibility of using a data commons model that allows branching. Nevertheless, there are mechanisms inherent in the process that may promote sharing. Any stakeholder wishing to receive new information from the data commons will need to address conflicts on each synchronization, but unless those resolved conflicts are fixed upstream, they will need to be addressed every time the organization needs to synchronize. Thus, the inherent value of the data in the commons promotes stakeholder-vetted contributions.

We recommend that the data commons impose governance functions that distribute the burden of many data integration tasks among community members. The data commons provides a centralized arbiter that would be readily available for access and queries at any time and could provide some guarantees on fast access to accurate and updated data. Importantly, rather than becoming a bulk repository for raw data, a community-mediated data commons could impose standardized structure upon data entry. Hence, the onus of data cleaning and validation is distributed both among stakeholders and temporally, in incremental updates.

12.7 FUTURE WORK

Because planning trips for pedestrians with disabilities has been limited by data availability, there remain many unanswered user experience questions. For example, AccessMap currently relies on the traffic light metaphor (green, yellow, red) to communicate elevation change, a convention that is not effective for all cultures and does not serve many colorblind users. In addition, much of the data used by AccessMap, such as sidewalk and crossing locations, could be applied to provide trips for people with visual impairments, a user class whose needs are currently not addressed by this application. Visualization and data consumption questions provide fertile ground for future work.

Existing and emerging technologies promote the large-scale, verifiable gathering data passively by non-experts, particularly pedestrian geodata. For example, smartphones can leverage GPS data, magnetometers, gyroscopes, and accelerometers to measure the incline of sidewalks (steepness at a location with a particular heading), with minimal input by the user. Phone GPSs combined with a series of images can reconstruct 3D models using structure from motion software. Finally, advances in low-cost lidar technology and increasingly common virtual reality applications could soon mean that high-resolution 3D scanning technology becomes ubiquitous in urban areas, making it possible to carry out more detailed surveys of the built environment by non-experts.

CHAPTER 13. A TECHNICAL OVERVIEW OF THE IMPLEMENTATION OF ACCESSMAP

This section serves as a reference for the implementation of AccessMap as of the time of writing. This reference may be used to reproduce the work of AccessMap as well as to reach a holistic understanding of the elements required to build a personalizable accessibility map at scale.

13.1 OVERVIEW

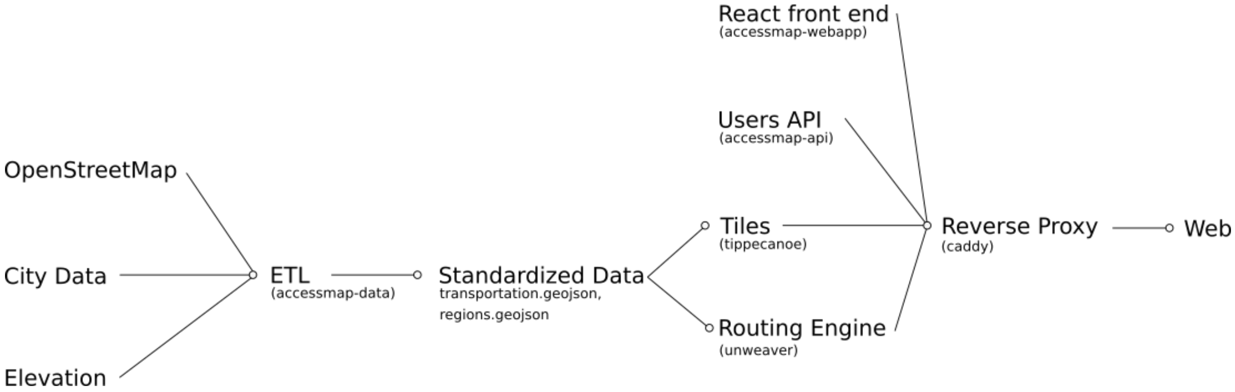


Figure 13.1: **The flow of information within the implementation of AccessMap**

The information displayed on AccessMap can be visualized as a directed acyclic graph (DAG) with either data pipelines or software as nodes and data as edges. Raw data from agencies or OpenStreetMap is processed via a reproducible ETL into the OpenSidewalks schema data standard. This data is then used to produce vector tiles and a routing engine and displayed via a single page application framework (React). Security and authentication are handled by a custom users API and a reverse proxy (Caddy).

AccessMap is implemented as a single page application (SPA) using the React framework with a Redux state store (Figure 13.1). Services and data are implemented as separate web APIs (users API, tiles API, the routing engine) and tied together using a reverse proxy (Caddy) that implements automatic HTTPS capabilities through LetsEncrypt. Both the tile server and the routing engine consume data in the OpenSidewalks schema format, commonly exchanged in a

GeoJSON format as a “transportation.geojson”. file These data are created through disparate ETL pipelines that combine together agency, OpenStreetMap, and elevation data.

13.2 DATA FLOWS

13.2.1 *Seattle, WA*

Figure 13.2: **The AccessMap Seattle data flow**

The DAG for AccessMap Seattle. Municipal data sources are cleaned and joined following practices described in Chapter 6.

The Seattle workflow (Figure 13.2) is relatively complex, requiring the combination of six datasets, including elevation data. Data are combined and extracted using modern versions of the algorithms described in Chapter 6.

13.2.2 *Mount Vernon, WA*

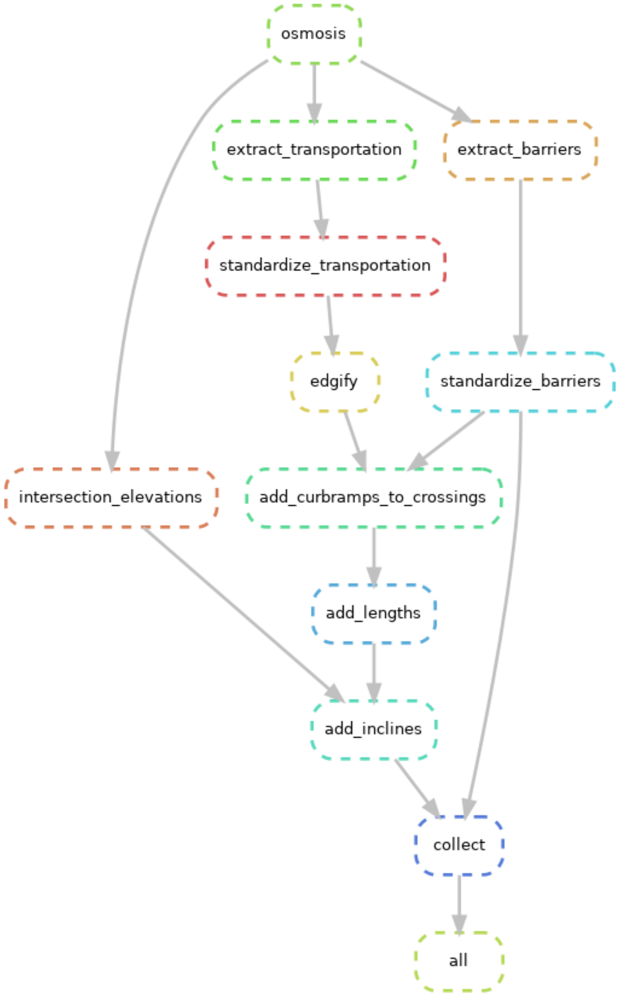


Figure 13.3: **The AccessMap Mount Vernon data flow**

The DAG for AccessMap Mount Vernon. The Mount Vernon workflow derives its vector (network) data exclusively from OpenStreetMap.

The Mount Vernon data workflow is substantially simpler than that of Seattle, as it derives all of its vector data from OpenStreetMap in a format that is highly similar to the OpenSidewalks schema (Figure 13.3). Nearly every step is focused on minor attribute transformations to produce data in the OpenSidewalks schema format or in deriving reasonable path segments from OSM data, which does not split data at intersections by default ("edgify").

13.2.3 *Bellingham, WA*

Figure 13.4: **The AccessMap Bellingham data flow**

The DAG for AccessMap Bellingham. The Bellingham workflow derives its data from municipal sources like Seattle, but in different formats and with different conventions.

The data workflow for Bellingham is similarly complex as Seattle and required the creation of custom data-stitching logic. Bellingham GIS provides highly-accurate sidewalk centerlines and crossing locations, but they are not split in a way that is amenable to the creation of a pedestrian network: paths extend around corners at intersects, for example. Most of the data processing is spent on estimating reasonable locations to split sidewalk polylines based on spatial relationships.

13.3 A DEEPLY INTERACTIVE MAP: VECTOR TILES, WEBGL MAPS, AND A SINGLE-PAGE-APPLICATION IMPLEMENTATION

Flexibility and individualized feedback were core goals of AccessMap, as it must meet diverse informational needs, be amenable to research purposes, and provide as much feedback and information as possible. AccessMap is therefore designed with a highly-interactive map based on vector tiles. In traditional web mapping applications, square images are pre-rendered from data for different zoom levels and any deeper investigation of the underlying data would depend on a distinct API request: clicking on a section of the map, sending a request to a backend server, and receiving information. If this information were displayed on the map, it would need to be fetched separately and custom-styled, often using DOM SVG elements with practical limitations on scaling: approximately 1,000 SVG elements can lead to slowdowns on web browsers, particularly on mobile phones. Vector tiles allow a more direct means of interacting with map data: the data itself is cut into square regions that are then recombined in the browser and rendered directly by a compositor. For a website like AccessMap, this is WebGL-based using the Mapbox-GL-JS library[88].

Because the data is directly embedded in the map, this allows for unique and highly-responsive modes of interaction in AccessMap: the data can trigger immediate re-renders of map elements based on user preferences or queries and map data can be investigated directly through mouse clicks that produce pop-ups without a round trip to the server. This also means there is a direct and up-to-date correspondence between these map data and the routing server, as they derive from the same OpenSidewalks schema-derived data.

AccessMap's sidewalks are differentially rendered using a styling DSL defined by Mapbox, the Mapbox style spec[89]. Sidewalks that are outside of a user's preferences are styled as dotted red lines while those within them are colored according to the cost function used in the backend, but reimplemented in JavaScript, implementing Tobler's hiking function to estimate difficulty of travel. This is then translated through a "traffic light" color scheme which, while less than ideal for color-blind users, is easy for users to understand (future work will enable a color-blind mode). Consequently, as AccessMap users dial in their personalized settings for uphill and downhill steepness, the map changes in real-time, providing instant feedback on their choices so that they can orient their personal preferences versus map data. A similar approach is taken for other pedestrian pathways, including crossings: crossings that are inaccessible according to a user’s preferences, such as crossings lacking curb ramps for a user that requires them, as colored as inaccessible via dotted red lines. Accessible crossings are then colored based on whether they are marked crossings or not. By toggling the “avoid raised curbs” switch, users will see crossings become available or unavailable in real-time.

13.4 PERSONAL MAPS: SETTINGS, LOGINS, AND PROFILE STORAGE

Pedestrian routing profiles are highly personal, suggesting a benefit to saving settings on a per-individual basis. Once AccessMap users have dialed in their personal settings, they have the option of creating an account and saving those preferences. At the time of writing, this process is mediated by the AccessMap-API[73] codebase, which mediates between OAuth 1.0a and OAuth2 portals and a database backend to securely store and retrieve pedestrian profile information. This database may eventually serve as a valuable resource for research and pedestrian routing, particularly in combination with demographic information, as part of the future directions research suggested in Chapter 14.

CHAPTER 14. DISCUSSION

14.1 CULTURAL AND GEOGRAPHIC LIMITATIONS IN SCOPE

AccessMap and related work have been designed around the concerns, geography, and cultural considerations of the United States, particularly Washington State. While many of the paradigms and pedestrian concerns are likely to translate into other geographies and cultures, it is unclear how well they will be directly convertible in a wildly different context. For example, the primary pedestrian ways in many urban spaces around the world are compacted dirt roads or otherwise do not have a discrete, protected area to separate foot traffic from car traffic from animal traffic. In such a situation, SidewalkScore would fail to have immediate meaning and AccessMap could only reasonably operate on streets and potential barriers along them or their peripheries.

14.2 IDENTIFICATION OF PEDESTRIAN MOBILITY PREFERENCE CLUSTERS

Our research in Chapter 4 regarding pedestrian preferences has several limitations. While the sample size is comparable to or larger than several past studies, it is still too small to confidently extract certain statistics or conclusions that a larger dataset would enable, such as natural clusters of preference within or between disability populations. For example, there may be five or more distinct populations of manual wheelchair users with correlated mobility preferences, but our study only sampled four individuals. In addition, some mobility preferences may be difficult to capture through self-assessments. One segment of our study (data not shown) asked for self-assessments of athleticism and strength, but survey results were inconsistent with impressions provided at in-person interviews. For example, two manual wheelchair users that would regularly jump curbs both stated that they exercised regularly and indicated higher propensity to perform athletic feats when traversing the environment, but gave themselves average or below average athleticism scores. Deeper research into the phrasing of questions could be measured against objective metrics of accessibility, establishing means by which to reliably scale self-assessment accessibility studies. In addition, the ability to voluntarily track pedestrian GPS locations could shed light on behaviors and preferences without the need for self-assessment, such as decisions to avoid hills. Finally, there are many disability populations that our research did not sufficiently probe, including individuals who use prosthetics or crutches, have a mobility limitation but do not use an assistive device, have a vision or hearing impairment, or have multiple intersecting disabilities or non-disability intersections of importance. Future work that addresses these limitations may reveal otherwise unrecognized disability subpopulations that can be served by technological or computing interventions, improving access to navigation information and accessibility audits.

Furthermore, exploration of pedestrian mobility preferences provide an opportunity for clustering or coincidence analysis that may challenge stereotypes regarding pedestrian needs. For example, it is clear that manual wheelchair users do not have monolithic preferences: there are at least two subgroups within the manual wheelchair population that have differing preferences. Larger-scale studies into pedestrian behavior and stated preferences could be analyzed to naturally discover new and more valid clusters of pedestrian interest that could be used to inform informational tools like AccessMap, analyses like SidewalkScore, and policy.

14.3 ALTERNATIVE "BEST PATH" DISCOVERY IMPLEMENTATIONS

While a routable pedestrian network is necessary for modern approaches to efficient "best path" discovery, there are many alternatives that could be explored going forward.

The first is the traversal algorithms for discovering best paths and their limitations. Dijkstra and similar approaches are, as mentioned in Chapter 10.1.1, limited to evaluating edges independent from one another in order to derive weights, which are then summed over paths. This consideration will fail once a fatigue model is implemented, as fatigue necessarily uses the history of the path traveled in order to calculate new weights, whereas the optimality of Dijkstra's method depends on these being independent. For example, a steep path at the end of a trip may be more difficult than one at the beginning. In addition, as we split up edges to account for greater specificity and variation, such as accounting for surface material changes, we lose the ability to easily account for aggregate effects, such as it being more difficult to traverse a long uphill path than a series of short ones: these could be considered identical in a Dijkstra-based approach. Modifications to Dijkstra, or the use of entirely different graph traversal algorithms that do use history, such as time-dependent fastest path algorithms[17].

The second is the way in which pedestrian networks are derived. Two-dimensional (or even three-dimensional) representations of pedestrian spaces may be more appropriate as ground truth data, both because of their increased accuracy and expressiveness of the problem of pedestrian traversability and because methods for collecting such data are rapidly expanding, including planimetric orthoimagery and surface density measurement, LIDAR, and Structure from Motion and photogrammetry techniques. Pedestrian networks may, in the future be more appropriately described as derivations of such data, opening opportunities to analyze a wider range of potential pedestrian paths. For example, a robotics pathfinding algorithms may discover more appropriate paths for wheelchair users with different wheelchair widths and lengths than could be reasonably mapped with manually-curated path width information. Such algorithms could be put to immediate use in many current real-world cases, as pedestrian plazas are already recorded as the equivalent of MultiPolygons in OpenStreetMap but routing engines lack the functionality to extract pedestrian paths within them.

BIBLIOGRAPHY

[1] Ali Abdolrahmani, William Easley, Michele A. Williams, Erick Ronquillo, Stacy Branham, Tiffany Chen, and Amy Hurst. 2016. Not All Errors are Created Equal: Factors that Impact Acceptance of an Indoor Navigation Aid for the Blind. In *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility* (ASSETS '16), Association for Computing Machinery, Reno, Nevada, USA, 301–302. DOI:https://doi.org/10.1145/2982142.2982210

[2] Adriano Alessandrini, Andrea Campagna, Paolo Delle Site, Francesco Filippi, and Luca Persia. 2015. Automated Vehicles and the Rethinking of Mobility and Cities. *Transp. Res. Procedia* 5, (2015), 145–160. DOI:https://doi.org/10.1016/j.trpro.2015.01.002

[3] Oladele A. Atoyebi, Delphine Labbé, Mike Prescott, Atiya Mahmood, François Routhier, William C. Miller, and W. Ben Mortenson. 2019. Mobility Challenges Among Older Adult Mobility Device Users. *Curr. Geriatr. Rep.* 8, 3 (September 2019), 223–231. DOI:https://doi.org/10.1007/s13670-019-00295-5

[4] Shiri Azenkot, Sanjana Prasain, Alan Borning, Emily Fortuna, Richard E. Ladner, and Jacob O. Wobbrock. 2011. Enhancing independence and safety for blind and deaf-blind public transit riders. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '11), Association for Computing Machinery, Vancouver, BC, Canada, 3247–3256. DOI:https://doi.org/10.1145/1978942.1979424

[5] Linda Beale, Kenneth Field, David Briggs, Phil Picton, and Hugh Matthews. 2006. Mapping for Wheelchair Users: Route Navigation in Urban Spaces. *Cartogr. J.* 43, 1 (March 2006), 68–81. DOI:https://doi.org/10.1179/000870406X93517

[6] Greg Bloom. 2013. Towards a community data commons. *Transpar. Open Data Future Civ. Innov.* (2013), 255–270.

[7] N. Bolten, S. Mukherjee, V. Sipeeva, A. Tanweer, and A. Caspi. 2017. A pedestrian-centered data approach for equitable access to urban infrastructure environments. *IBM J Res Dev* (2017). DOI:https://doi.org/10.1147/JRD.2017.2736279

[8] Stacy M. Branham, Ali Abdolrahmani, William Easley, Morgan Scheuerman, Erick Ronquillo, and Amy Hurst. 2017. "Is Someone There? Do They Have a Gun": How Visual Information about Others Can Improve Personal Safety Management for Blind Individuals. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility* (ASSETS '17), Association for Computing Machinery, Baltimore, Maryland, USA, 260–269. DOI:https://doi.org/10.1145/3132525.3132534

[9] M Brewster, D Hurtado, Sara Olson, and Jessica Yen. 2009. Walkscore.com: A new methodology to explore associations between neighborhood resources, race, and health. *APHA Poster Present.* (2009). Retrieved from https://apha.confex.com/recording/apha/137am/pdf/free/4db77adf5df9fff0d3caf5cafe28f496/paper205082_1.pdf

[10] Megan Campbell, Cynthia Bennett, Caitlin Bonnar, and Alan Borning. 2014. Where's my bus stop? supporting independence of blind transit riders with StopInfo. In *Proceedings of the 16th international ACM SIGACCESS conference on Computers & accessibility* (ASSETS

’14), Association for Computing Machinery, Rochester, New York, USA, 11–18. DOI:https://doi.org/10.1145/2661334.2661378

[11] Cameron Tyler Cassidy. 2017. Identifying Visual Cues to Improve Independent IndoorNavigation for Blind Individuals. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility* (ASSETS ’17), Association for Computing Machinery, Baltimore, Maryland, USA, 413–414. DOI:https://doi.org/10.1145/3132525.3134828

[12] Yao-Jen Chang, Shu-Ming Peng, Tsen-Yung Wang, Shu-Fang Chen, Yan-Ru Chen, and Hung-Chi Chen. 2010. Autonomous indoor wayfinding for individuals with cognitive impairments. *J. NeuroEngineering Rehabil.* 7, 1 (September 2010), 45. DOI:https://doi.org/10.1186/1743-0003-7-45

[13] Mandy Chessell, Ferd Scheepers, Nhan Nguyen, Ruud van Kessel, and Ron van der Starre. 2014. *Governing and Managing Big Data for Analytics and Decision Makers*. IBM.

[14] Open Geospatial Consortium and others. 2015. OGC GeoPackage Encoding Standard with corrigendum. 10, 12 (2015), 2015.

[15] Jorge L. Contreras. 2014. Constructing the Genome Commons. In *Governing Knowledge Commons*, Brett M. Frischmann, Michael J. Madison and Katherine J. Strandburg (eds.). Oxford University Press, 99–136. DOI:https://doi.org/10.1093/acprof:oso/9780199972036.003.0004

[16] Rebecca C. Deehr and Amy Shumann. 2009. Active Seattle: achieving walkability in diverse neighborhoods. *Am. J. Prev. Med.* 37, 6 (2009), S403–S411.

[17] Ugur Demiryurek, Farnoush Banaei-Kashani, Cyrus Shahabi, and Anand Ranganathan. 2011. Online Computation of Fastest Path in Time-Dependent Spatial Networks. In *Advances in Spatial and Temporal Databases*, Dieter Pfoser, Yufei Tao, Kyriakos Mouratidis, Mario A. Nascimento, Mohamed Mokbel, Shashi Shekhar and Yan Huang (eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 92–111. DOI:https://doi.org/10.1007/978-3-642-22922-0_7

[18] Dan Ding, Bambang Parmanto, Hassan A. Karimi, Duangduen Roongpiboonsopit, Gede Pramana, Thomas Conahan, and Piyawan Kasemsuppakorn. 2007. Design Considerations for a Personalized Wheelchair Navigation System. In *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 4790–4793. DOI:https://doi.org/10.1109/IEMBS.2007.4353411

[19] Anhai Doan, Raghu Ramakrishnan, and Alon Y. Halevy. 2011. Crowdsourcing systems on the World-Wide Web. *Commun. ACM* 54, 4 (April 2011), 86–96. DOI:https://doi.org/10.1145/1924421.1924442

[20] John Robert Edmonds Jr. 1960. A combinatorial representation for oriented polyhedral surfaces. PhD Thesis.

[21] Geraint Ellis, Ruth Hunter, Mark A Tully, Michael Donnelly, Luke Kelleher, and Frank Kee. 2016. Connectivity and physical activity: using footpath networks to measure the walkability of built environments. *Environ. Plan. B Plan. Des.* 43, 1 (January 2016), 130–151. DOI:https://doi.org/10.1177/0265813515610672

[22] Brian Ferris, Kari Watkins, and Alan Borning. 2010. OneBusAway: results from providing real-time arrival information for public transit. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI ’10), Association for Computing Machinery, Atlanta, Georgia, USA, 1807–1816. DOI:https://doi.org/10.1145/1753326.1753597

[23] Seth Fitzsimmons. 2016. Merging Offline Edits with the POSM Replay Tool. *Medium*. Retrieved June 26, 2020 from https://hi.stamen.com/merging-offline-edits-with-the-posm-replay-tool-2f39a4410d2a

[24] Brett M. Frischmann, Michael J. Madison, and Katherine Jo Strandburg (Eds.). 2014. *Governing knowledge commons*. Oxford University Press, Oxford ; New York.

[25] Jon E. Froehlich, Anke M. Brock, Anat Caspi, João Guerreiro, Kotaro Hara, Reuben Kirkham, Johannes Schöning, and Benjamin Tannert. 2019. Grand challenges in accessible maps. Retrieved May 7, 2020 from https://doi.org/10.1145/3301657

[26] Jonathan M. Gallimore, Barbara B. Brown, and Carol M. Werner. 2011. Walking routes to school in new urban and suburban neighborhoods: An environmental walkability analysis of blocks and routes. *J. Environ. Psychol.* 31, 2 (June 2011), 184–191. DOI:https://doi.org/10.1016/j.jenvp.2011.01.001

[27] Nicholas A. Giudice and Gordon E. Legge. 2008. Blind Navigation and the Role of Technology. In *The Engineering Handbook of Smart Technology for Aging, Disability, and Independence*. John Wiley & Sons, Ltd, 479–500. DOI:https://doi.org/10.1002/9780470379424.ch25

[28] Klokan Technologies GmbH. OpenMapTiles Vector Tile Schema. *OpenMapTiles*. Retrieved June 25, 2020 from https://openmaptiles.org/schema/

[29] Maya Gupta, Ali Abdolrahmani, Emory Edwards, Mayra Cortez, Andrew Tumang, Yasmin Majali, Marc Lazaga, Samhitha Tarra, Prasad Patil, Ravi Kuber, and Stacy M. Branham. 2020. Towards More Universal Wayfinding Technologies: Navigation Preferences Across Disabilities. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (CHI ’20), Association for Computing Machinery, Honolulu, HI, USA, 1–13. DOI:https://doi.org/10.1145/3313831.3376581

[30] Laura M. Haas, Mauricio A. Hernández, Howard Ho, Lucian Popa, and Mary Roth. 2005. Clio grows up: from research prototype to industrial tool. In *Proceedings of the 2005 ACM SIGMOD international conference on Management of data - SIGMOD ’05*, ACM Press, Baltimore, Maryland, 805. DOI:https://doi.org/10.1145/1066157.1066252

[31] Aric Hagberg, Pieter Swart, and Daniel S Chult. 2008. Exploring network structure, dynamics, and function using networkx. (2008).

[32] Alon Halevy, Anand Rajaraman, and Joann Ordille. 2006. *Data Integration: The Teenage Years.*

[33] Kotaro Hara, Shiri Azenkot, Megan Campbell, Cynthia L. Bennett, Vicki Le, Sean Pannella, Robert Moore, Kelly Minckler, Rochelle H. Ng, and Jon E. Froehlich. 2015. Improving Public Transit Accessibility for Blind Riders by Crowdsourcing Bus Stop Landmark Locations with Google Street View: An Extended Analysis. Retrieved May 7, 2020 from https://doi.org/10.1145/2717513

[34] Kotaro Hara, Christine Chan, and Jon E. Froehlich. 2016. The Design of Assistive Location-based Technologies for People with Ambulatory Disabilities: A Formative Study. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (CHI ’16), Association for Computing Machinery, San Jose, California, USA, 1757–1768. DOI:https://doi.org/10.1145/2858036.2858315

[35] Mahdi Hashemi and Hassan A. Karimi. 2017. Collaborative personalized multi-criteria wayfinding for wheelchair users in outdoors. *Trans. GIS* 21, 4 (August 2017), 782–795. DOI:https://doi.org/10.1111/tgis.12230

[36] D Richard Hipp. 2000. SQLite. (2000). Retrieved from http://sqlite.org/

[37] Yoshie Inada, Shinobu Izumi, Motoya Koga, and Shigehito Matsubara. 2014. Development of Planning Support System for Welfare Urban Design – Optimal Route Finding for Wheelchair Users. *Procedia Environ. Sci.* 22, (January 2014), 61–69. DOI:https://doi.org/10.1016/j.proenv.2014.11.006

[38] Hans J. (Jochen) Scholl and Ralf Klischewski. 2007. E-Government Integration and Interoperability: Framing the Research Agenda. *Int. J. Public Adm.* 30, 8–9 (July 2007), 889–920. DOI:https://doi.org/10.1080/01900690701402668

[39] Bumjoon Kang, Jason Y. Scully, Orion Stewart, Philip M. Hurvitz, and Anne V. Moudon. 2015. Split-Match-Aggregate (SMA) algorithm: integrating sidewalk data with transportation network data in GIS. *Int. J. Geogr. Inf. Sci.* 29, 3 (March 2015), 440–453. DOI:https://doi.org/10.1080/13658816.2014.981191

[40] Peter Karich and Stefan Schröder. 2014. Graphhopper. 4, 2 (2014), 15.

[41] Hassan A. Karimi, Lei Zhang, and Jessica G. Benner. 2014. Personalized accessibility map (PAM): A novel assisted wayfinding approach for people with disabilities. *Ann. GIS* 20, 2 (2014), 99–108.

[42] Piyawan Kasemsuppakorn, Hassan A. Karimi, Dan Ding, and Manoela A. Ojeda. 2015. Understanding route choices for wheelchair navigation. *Disabil. Rehabil. Assist. Technol.* 10, 3 (May 2015), 198–210. DOI:https://doi.org/10.3109/17483107.2014.898160

[43] Rob Kitchin. 2014. *The data revolution: big data, open data, data infrastructures & their consequences*. SAGE Publications, Los Angeles, California.

[44] Rainer Kuhlen. 2012. Knowledge is the water of the mind : how to structure rights in immaterial commons. In *The wealth of the commons : a world beyond market and state*, David Bollier and Silke Helfrich (eds.). Levellers, Amherst, Mass., 331–339.

[45] Anthony Li, Manaswi Saha, Anupam Gupta, and Jon E. Froehlich. 2018. Interactively Modeling and Visualizing Neighborhood Accessibility at Scale: An Initial Study of Washington DC. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility - ASSETS '18*, ACM Press, Galway, Ireland, 444–446. DOI:https://doi.org/10.1145/3234695.3241000

[46] B McHugh. 2011. The OpenTripPlanner project. *OpenTripPlanner Proj.* (2011), 12–16.

[47] Christian Menkens, Julian Sussmann, Miriam Al-Ali, Emanuel Breitsameter, Jelena Frtunik, Tobias Nendel, and Tobias Schneiderbauer. 2011. EasyWheel - A Mobile Social Navigation and Support System for Wheelchair Users. In *2011 Eighth International Conference on Information Technology: New Generations*, 859–866. DOI:https://doi.org/10.1109/ITNG.2011.149

[48] Amin Mobasheri, Jonas Deister, and Holger Dieterich. 2017. Wheelmap: the wheelchair accessibility crowdsourcing platform. *Open Geospatial Data Softw. Stand.* 2, 1 (November 2017), 27. DOI:https://doi.org/10.1186/s40965-017-0040-5

[49] Humanitarian OpenStreetMapTeam. Tasking Manager. *Tasking Manager*. Retrieved June 25, 2020 from https://tasks.hotosm.org

[50] Elinor Ostrom. 1990. *Governing the commons: the evolution of institutions for collective action*. Cambridge University Press, Cambridge ; New York.

[51] Ben Pierce, Eric Plapper, and Jodi Rizek. 2016. Accessible Transportation Technologies Research Initiative (ATTRI) : User Needs Assessment: Stakeholder Engagement Report. FHWA-JPO-16-354 (May 2016). Retrieved from https://rosap.ntl.bts.gov/view/dot/31320

[52] L. Ran, S. Helal, and S. Moore. 2004. Drishti: an integrated indoor/outdoor blind navigation system and service. In *Second IEEE Annual Conference on Pervasive Computing*

and Communications, 2004. Proceedings of the, 23–30. DOI:https://doi.org/10.1109/PERCOM.2004.1276842

[53] Alejandro Rituerto, Giovanni Fusco, and James M. Coughlan. 2016. Towards a Sign-Based Indoor Navigation System for People with Visual Impairments. In *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility* (ASSETS ’16), Association for Computing Machinery, Reno, Nevada, USA, 287–288. DOI:https://doi.org/10.1145/2982142.2982202

[54] Philipp Rode, Graham Floater, Nikolas Thomopoulos, James Docherty, Peter Schwinger, Anjali Mahendra, and Wanli Fang. 2017. Accessibility in Cities: Transport and Urban Form. In *Disrupting Mobility: Impacts of Sharing Economy and Innovative Transportation on Cities*, Gereon Meyer and Susan Shaheen (eds.). Springer International Publishing, Cham, 239–273. DOI:https://doi.org/10.1007/978-3-319-51602-8_15

[55] Carol Rose. 1986. The Comedy of the Commons: Custom, Commerce, and Inherently Public Property. *Univ. Chic. Law Rev.* 53, 3 (1986), 711. DOI:https://doi.org/10.2307/1599583

[56] Paula W. Rushton, Ben W. Mortenson, Pooja Viswanathan, Rosalie H. Wang, William C. Miller, Laura Hurd Clarke, and CanWheel Research Team. 2017. Intelligent power wheelchair use in long-term care: potential users’ experiences and perceptions. *Disabil. Rehabil. Assist. Technol.* 12, 7 (October 2017), 740–746. DOI:https://doi.org/10.1080/17483107.2016.1260653

[57] Jaime Sánchez and Natalia de la Torre. 2010. Autonomous navigation through the city for the blind. In *Proceedings of the 12th international ACM SIGACCESS conference on Computers and accessibility* (ASSETS ’10), Association for Computing Machinery, Orlando, Florida, USA, 195–202. DOI:https://doi.org/10.1145/1878803.1878838

[58] Harold Shields, Sindre Fredsvik, Jonah Horn, Robert Kalei Miller, and Cameron Hill. *Portland Region Parks: Measuring Equity in Access*. Retrieved from http://pdxscholar.library.pdx.edu/cgi/viewcontent.cgi?article=1013&context=ims_assestmapping

[59] Mapillary AB Sweden. The street-level imagery platform that scales and automates mapping. *Mapillary*. Retrieved June 25, 2020 from https://mapillary.com

[60] Ignacio G. Terrizzano, Peter M. Schwarz, Mary Roth, and John E. Colino. 2015. Data Wrangling: The Challenging Yourney from the Wild to the Lake. In *CIDR*.

[61] Nagendra R. Velaga, Mark Beecroft, John D. Nelson, David Corsar, and Peter Edwards. 2012. Transport poverty meets the digital divide: accessibility and connectivity in rural communities. *J. Transp. Geogr.* 21, (March 2012), 102–112. DOI:https://doi.org/10.1016/j.jtrangeo.2011.12.005

[62] Pooja Viswanathan, James J Little, Alan K Mackworth, Tuck-Voon How, Rosalie H Wang, and Alex Mihailidis. 2013. Intelligent wheelchairs for cognitively-impaired older adults in Long-term care: A review. In *Proceedings of the Rehabilitation Engineering and Assistive Technology Society of North America Annual Conference*.

[63] Thorsten Völkel and Gerhard Weber. 2008. RouteCheckr: personalized multicriteria routing for mobility impaired pedestrians. In *Proceedings of the 10th international ACM SIGACCESS conference on Computers and accessibility - Assets ’08*, ACM Press, Halifax, Nova Scotia, Canada, 185. DOI:https://doi.org/10.1145/1414471.1414506

[64] Lori Walters. 2016. Small Commercial Data Pool Project— Leveraging Partnership. Retrieved from https://media.genre.com/documents/iinapc1611-en.pdf

[65] Kari Edison Watkins, Brian Ferris, Alan Borning, G. Scott Rutherford, and David Layton. 2011. Where Is My Bus? Impact of mobile real-time information on the perceived and actual wait time of transit riders. *Transp. Res. Part Policy Pract.* 45, 8 (October 2011), 839–848. DOI:https://doi.org/10.1016/j.tra.2011.06.010

[66] Michele A. Williams, Caroline Galbraith, Shaun K. Kane, and Amy Hurst. 2014. “just let the cane hit it”: how the blind and sighted see navigation differently. In *Proceedings of the 16th international ACM SIGACCESS conference on Computers & accessibility* (ASSETS ’14), Association for Computing Machinery, Rochester, New York, USA, 217–224. DOI:https://doi.org/10.1145/2661334.2661380

[67] Michele A. Williams, Amy Hurst, and Shaun K. Kane. 2013. “Pray before you step out”: describing personal and situational blind navigation behaviors. In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility* (ASSETS ’13), Association for Computing Machinery, Bellevue, Washington, 1–8. DOI:https://doi.org/10.1145/2513383.2513449

[68] Rayoung Yang, Sangmi Park, Sonali R. Mishra, Zhenan Hong, Clint Newsom, Hyeon Joo, Erik Hofer, and Mark W. Newman. 2011. Supporting spatial awareness and independent wayfinding for pedestrians with visual impairments. In *The proceedings of the 13th international ACM SIGACCESS conference on Computers and accessibility* (ASSETS ’11), Association for Computing Machinery, Dundee, Scotland, UK, 27–34. DOI:https://doi.org/10.1145/2049536.2049544

[69] Limin Zeng, Denise Prescher, and Gerhard Weber. 2012. Exploration and avoidance of surrounding obstacles for the visually impaired. In *Proceedings of the 14th international ACM SIGACCESS conference on Computers and accessibility* (ASSETS ’12), Association for Computing Machinery, Boulder, Colorado, USA, 111–118. DOI:https://doi.org/10.1145/2384916.2384936

[70] Tobias Zwick. 2020. *westnordost/StreetComplete*. Retrieved June 25, 2020 from https://github.com/westnordost/StreetComplete

[71] 1993. ADA, Disability Rights, and Evolving Regulatory Federalism. *Publius J. Fed.* (1993). DOI:https://doi.org/10.1093/oxfordjournals.pubjof.a038104

[72] 2018. Introducing “wheelchair accessible” routes in transit navigation. *Google*. Retrieved June 24, 2020 from https://blog.google/products/maps/introducing-wheelchair-accessible-routes-transit-navigation/

[73] 2019. *AccessMap/accessmap-api*. AccessMap. Retrieved June 25, 2020 from https://github.com/AccessMap/accessmap-api

[74] 2020. *AccessMap/sidewalkify*. AccessMap. Retrieved June 25, 2020 from https://github.com/AccessMap/sidewalkify

[75] 2020. *OpenSidewalks/OpenSidewalks-Schema*. OpenSidewalks. Retrieved June 25, 2020 from https://github.com/OpenSidewalks/OpenSidewalks-Schema

[76] AccessMap. Retrieved June 26, 2020 from https://www.accessmap.io

[77] Openrouteservice. Retrieved May 7, 2020 from https://openrouteservice.org/

[78] AXS Map. Retrieved June 26, 2020 from https://axsmap.com/

[79] Microsoft Soundscape. *Microsoft Research*. Retrieved June 26, 2020 from https://www.microsoft.com/en-us/research/product/soundscape/

[80] pathVu | Home. Retrieved June 26, 2020 from http://www.pathvu.com/

[81] NextBus. Retrieved June 26, 2020 from https://www.nextbus.com

[82] OpenStreetMap. *OpenStreetMap*. Retrieved June 25, 2020 from https://www.openstreetmap.org/

[83] Humanitarian OpenStreetMap Team | Home. Retrieved June 25, 2020 from https://www.hotosm.org/

[84] Overpass Turbo. Retrieved June 25, 2020 from https://overpass-turbo.eu/

[85] Field Papers. Retrieved June 25, 2020 from http://fieldpapers.org/

[86] *Project OSRM*. Retrieved from http://project-osrm.org

[87] *Non-Motorized Connectivity Study*. Fehr & Piers. Retrieved from http://metro.kingcounty.gov/programs-projects/nmcs/pdf/4-nmcs-report-091214-analysis.pdf

[88] Introduction. *Mapbox*. Retrieved June 25, 2020 from https://www.mapbox.com/mapbox-gl-js/api/

[89] Style Specification | Mapbox GL JS | Mapbox. Retrieved June 25, 2020 from https://docs.mapbox.com/mapbox-gl-js/style-spec/

VITA

Nicholas Bolten is a first-generation bachelor's degree recipient born in Missoula, Montana. He earned a double degree in Biochemistry and Microbiology from the University of Washington, during which he joined the Klavins Lab, working first as an assistant, then as a full-time researcher, and finally beginning a PhD pursuing research. His early PhD career was focused on the field of synthetic biology, designing genetic circuits in yeast and publishing a first author cover article in the field's main journal, ACS Synthetic Biology, before becoming captivated by the potential for impact by addressing the informational needs of people with mobility limitations under the mentorship of Dr. Anat Caspi of the Taskar Center for Accessible Technology. There, he developed the work presented in this thesis, including founding of the AccessMap and OpenSidewalks projects. Nicholas lives in Seattle with his partner, their sibling, and too many pets, where he spends his free time cooking, running a self-hosted cloud, and spending time with family and friends.