Approximating Accessibility of Regions from Incomplete Volunteered Data

Hadi Asghari
Jakob Stolberg-Larsen
Theresa Züger
hadi.asghari@hiig.de
jakob.stolberg.larsen@hiig.de
zueger@hiig.de
Alexander von Humboldt Institute for Internet and Society
Berlin, Germany

ABSTRACT

Being informed about the accessibility of neighborhoods, cities, and regions can help persons with disabilities in making travel and daily decisions. This information can also be useful and a pushing factor for supportive public policies. While accessibility mapping initiatives, such as Wheelmap.org, have enjoyed tremendous success and scale, they are still far from exhaustive, and their coverage contains biases stemming from volunteer practices. With the aid of the framework of causal statistics, we suggest approaches to adjust for these biases, with the end goal of providing helpful approximations of overall accessibility in different European geographical regions.

CCS CONCEPTS

• Human-centered computing → Empirical studies in accessibility; **• Computing methodologies → Causal reasoning and diagnostics**.

KEYWORDS

Accessibility, Causality, Metric Design, Wheelmap

ACM Reference Format:
Hadi Asghari, Jakob Stolberg-Larsen, and Theresa Züger. 2022. Approximating Accessibility of Regions from Incomplete Volunteered Data. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts (CHI '22 Extended Abstracts), April 29-May 5, 2022, New Orleans, LA, USA.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3491101.3519706

1 INTRODUCTION

The United Nation's Convention on the Rights of Persons with Disabilities (PWD) asks signatories to:

> "take appropriate measures to ensure to persons with disabilities access, on an equal basis with others, to the physical environment, to transportation, [...] and to other facilities and services open or provided to the public, both in urban and in rural areas".[16]

Over 160 countries, including all member states of the European Union (EU), have signed this Convention. Recognizing the rights of PWDs is undoubtedly a historic achievement, but it is only part of the way; ultimately, these rights need to be enforced.[2] This means many buildings and facilities need to be made accessible, which requires awareness, resources, and effort.

Initiatives such as Wheelmap[1] play an important role in raising awareness. Started by the German non-profit organization Sozialhelden in 2010, Wheelmap is a volunteer effort to mark places that are wheelchair accessible on a map.[14] As of October 2021, the project had gathered information on the accessibility status of over one million public places worldwide. Volunteers mark places with different intentions. One is to share knowledge about wheelchair accessible places, so as to make it easier to find such places. This is helpful, for instance, when picking a place for socializing.[15] Mapping only accessible places is, however, not enough for locations that people *need* to visit, such as, schools, doctors, or government offices. In such cases, knowing about lack of access is also helpful for PWDs. Taking this idea one step further, knowledge about the accessibility of whole neighborhoods, cities, and regions is similarly important information for wheelchair users, disability activists, and city planners focused on social equality.

In this paper, we investigate how information about the accessibility of regions can be estimated from Wheelmap's vast dataset. The key technical challenge in this task is the problem of *missing information*, or *selection bias*, in the volunteered data. To demonstrate this point, we can compare the statistics for the state of Berlin, with around 63,000 public places, with that of the Flevoland region of the Netherlands, with approx. 6,500 public places. In Berlin, 23% of places are marked as fully wheelchair accessible, 9% as partially accessible, 13% as inaccessible, and 55% are unmarked or unknown. The ratio of fully or partially accessible points is thus about 71% of all marked points. In comparison, in Flevoland, less than 4% of places are marked; Among them, 91% are fully accessible. If we naively ignore the portion of a city covered, we might conclude that Flevoland is more accessible than Berlin, which may not be the case.

We can tackle this approximation challenge with the help of the causal inference literature[4, 10]. The causal graph depicted in

[1] https://wheelmap.org

CHI '22 Extended Abstracts, April 29-May 5, 2022, New Orleans, LA, USA

ACM ISBN 978-1-4503-9156-6/22/04.
https://doi.org/10.1145/3491101.3519706

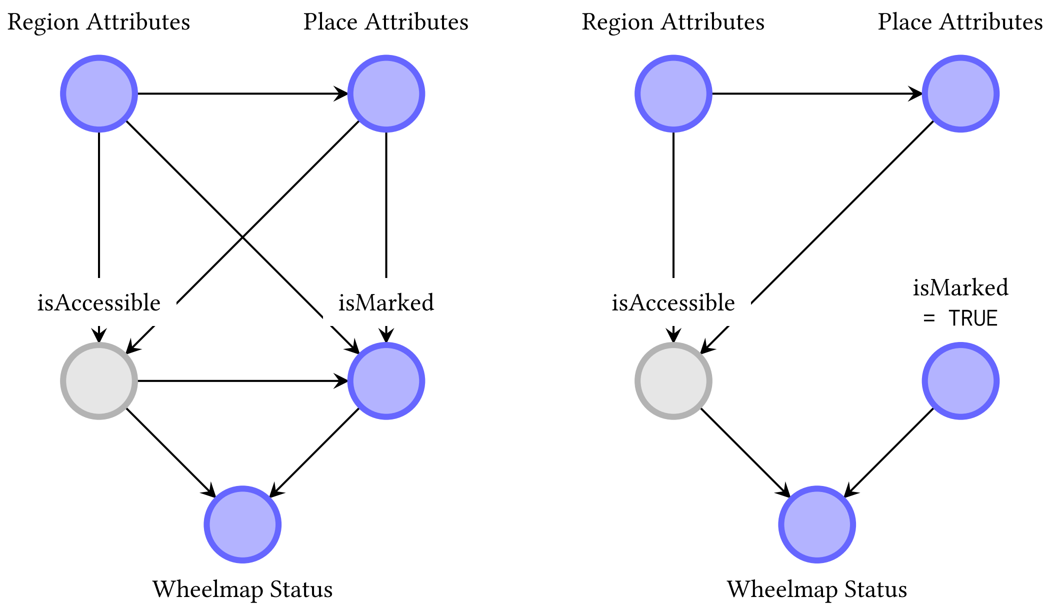

Figure 1: The left-hand causal graph models the relationship between a place's attributes and its region, its unobserved accessibility status, and its Wheelmap status. (Each node in the graph is a variable and the arrows represent causal relations). The right-hand graph depicts the same model after the intervention `do('isMarked'=TRUE)`; it is explained further in Section 3.

Fig.1-left models the high-level factors that lead to a place being marked in Wheelmap. The probability of a public place being accessible (or not) is influenced by the place attributes, for instance if it is a government building or a cafe, as well as by regional attributes, such as supportive policies, PWD community, urban density, infrastructure age, etc; whether a place is marked by volunteers is affected by similar factors —albeit with different effect sizes. The true accessibility of a place, which is an unobserved variable, and whether it is marked or not, combine into the place's status on Wheelmap. As we shall explain later in this paper, the link between the node `isAccessible` and the node `isMarked` is what causes the estimation problem. The framework of causality offers some solutions, including *do-operations*, which we shall also discuss.

This paper features work that is still in progress, hence its submission as a Late Breaking Work. As such, we are looking forward to receiving feedback from the CHI community regarding its two key contributions:

(1) Develop and articulate a method, using causal inference techniques, to make a better (unbiased) approximation of accessibility in an area or region, from vast yet incomplete volunteer data;
(2) Illustrate how an accessibility score for regions can be linked to various geographical and policy factors affecting access in those regions.

Given that the Wheelmap project and data has a European focus (since the community was started in Germany), we limit the analysis of the paper to European Union member states.

The rest of this paper is organized as follows: in Section 2 we describe the Wheelmap data; in Section 3 we present the causal inference framework and its use in our case; Section 4 offers a preliminary analysis; Section 5 looks at related work; and Section 6 concludes the paper.

2 WHEELMAP.ORG

The Wheelmap project is built on top of OpenStreetMap (OSM) data.[9] The accessibility data contributed by volunteers is typically stored back in OpenStreetMap as a series of tags.[2] The project includes a website and mobile apps.

The places to be marked are OSM *points of interest*, which broadly stated are public points or locations on a map that someone may find useful or interesting. These places include twelve major categories: public transport (such as bus stops and parking), food places (such as cafes and restaurants), leisure places (such as cinemas), money & post, education (including libraries), shopping places, sports centers, touristic attractions, hotels & accommodations, governmental buildings, healthcare facilities, and a miscellaneous category (which includes toilets and companies).

Places are marked according to their wheelchair accessibility based on a *traffic light system*:

- Green: fully wheelchair accessible—the entrance and all rooms are accessible without steps;
- Orange: partially wheelchair accessible—the entrance has one step (less than 7cm) and most important rooms are accessible without steps;
- Red: not wheelchair accessible—one or more steps and no temporary mobile ramp available.

[2]The exception is data contributed by partner organizations that do not have an open license and are consequently stored in a private cloud.

Places that are not yet marked with regards to their wheelchair accessibility are given a grey color. Wheelmap users are encouraged to mark them, and also upload photos of these places.

Due to links between our research group and Wheelmap.org, we were able to receive a full dump of the Wheelmap database in October 2021. The dump included a legacy MySQL table containing information on approximately 21 million places, plus some other structures and image files. Globally, 3.1% of all places are marked as accessible, 1% as partially accessible, 1.4% as inaccessible, bringing the total of marked places to 1.1 million; (thus 94.5% of places remain unmarked).

Each place has a geometric data field which points to a latitude and longitude. We use geodata from Natural Earth[3], as well as the EU's Nomenclature of territorial units for statistics (NUTS[4]) classification, to map these points to EU countries and regions. Over eighty percent all marked points are within the EU[5], which (as explained earlier) is why we limit the scope of the analysis to it.

3 METHODS

The topic of causality and causal inference has seen a revival in recent years, thanks in some part to the work of Judea Pearl and his colleagues[10, 11]. Pearl developed a calculus of causation, which makes use of *causal diagrams* and a symbolic machinery called the *do-calculus*, along with an accompanying set of theorems. We make use of these tools in this work as they aid us to better understand the data generation process and to think through possible remedies for the selection bias.

Causal diagrams, also known as direct acylic graphs (DAGs), are dot-and-arrow pictures that summarize our existing causal assumptions from prior knowledge. The dots represent random variables and are the quantities of interest, while "the arrows represent known or suspected causal relationships between those variables—namely, which variable 'listens' to which others"[11]. In Fig.1-left, we presented the simple DAG that we believe explains the status of locations in Wheelmap data. One benefit of using DAGs is that they make our assumptions explicit, allowing them to be investigated and discussed by others.

More importantly though, causal diagrams bring with them theorems, implications, and computational tools for calculating probabilities. Specifically, from our DAG we can conclude that the unobserved node `isAccessible` is not identifiable[1] (remains incalculable) due to the link between `isAccessible` and `isMarked`.[6] From the Wheelmap data, which can be regarded as observational data, we can estimate the total effect of `Region Attributes` & `Place Attributes` on `isMarked`, but not the direct effect of `isAccessible` on `isMarked`.

Stated differently, from our data, we can calculate `P(A|M=1)`, while what we are really interested in knowing is `P(A|do(M=1))`. The *do-operator* signifies that we are dealing with an 'intervention', which here is the act of marking the accessibility of an unknown place, rather than a passive observation (of seeing the accessibility of a *previously* marked place). In Pearl's words, "classical statistics has nothing remotely similar to the do operator"[11].

Herein also lies the possible remedy. In order to calculate the probabilities for `isAccessible` correctly (the true state of accessibility), we need a do-intervention on `isMarked`. A do-intervention on a node cuts all incoming links into that node—which here also severs the link between `isAccessible` and `isMarked`. Wheelmap hosts so called *mapathons*, or community mapping events, during which volunteers are invited to physically gather in a part of the city, and mark the accessibility of all places in that neighborhood.[17] Mapathons are effectively do-operations, and yield the graph on Fig.1-right (for the places marked during them). For places that are marked during a mapathon, we can calculate the unbiased access probability, and later use that probability to estimate other links in the graph.

Table 1: Regions with highest coverage (top) and highest access (bottom)

NUTS	Region Name	Coverage	Access	Places
DE30	Berlin	45%	70%	63,172
AL01	Veri	37%	93%	6,344
DE60	Hamburg	36%	80%	36,091
UKE1	E. Yorkshire & Northe	5%	97%	11,507
NL23	Flevoland	4%	96%	6,574
UKK2	Dorset & Somerset	3%	94%	15,749

4 PRELIMINARY ANALYSIS

4.1 Regions with highest and lowest map coverage and access

We can get a sense for the relationship between a place being accessible and being marked from simply listing the regions with the highest coverage (marked places) and the highest accessibility among the marked (Table 1). As can be seen, the regions with the highest access have very low coverage.

The coverage and access ratio of various European regions are depicted in the heatmaps of Fig.2. We can again see that the regions with the highest coverage have average access, and regions with the highest access have low coverage.

We can statistically test this relationship using *Spearman's rank correlation coefficient*. The correlation between the two ratios is -0.14 with a p value of 0.00; the correlation is even stronger (-0.24) if we limit it to regions with at least 100 points marked.[7] The negative correlation sign supports the existence of a link between `isAccessible` and `isMarked` in the DAG.

4.2 Detecting mapathons & other alternatives

As we have explained, one solution to correctly estimate accessibility from the Wheelmap data is to find places that are marked during a mapathon. As we have not yet found an exhaustive list of Wheelmap mapathons, we attempted two experiments to detect them computationally, based on temporal and spatial effects.

[3]https://www.naturalearthdata.com/
[4]https://ec.europa.eu/eurostat/web/nuts/background
[5]In particular, Germany and France have the most marked places, which reflects the origins of Wheelmap.org.
[6]As a reminder, this link states that accessible places are more likely to be initially marked by volunteers.
[7]The median number of pints marked in a region is 858.

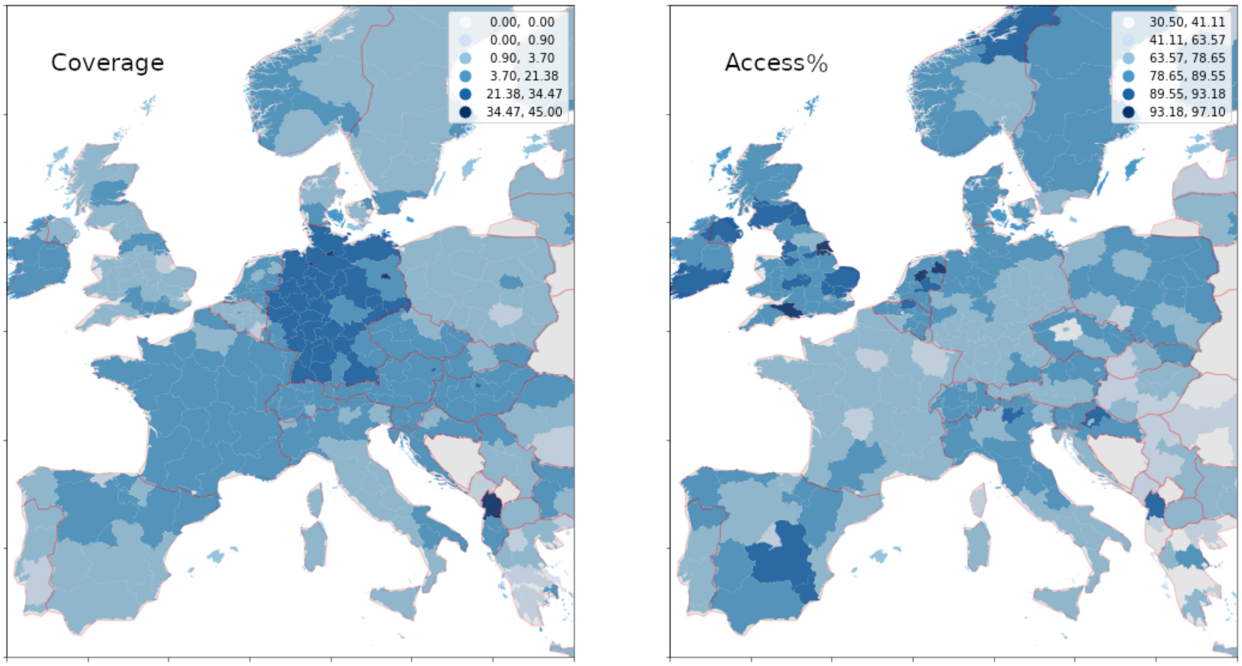

Figure 2: Heatmap of the Wheelmap coverage of NUTS2 regions (left); and the access ratio among the marked places (right)

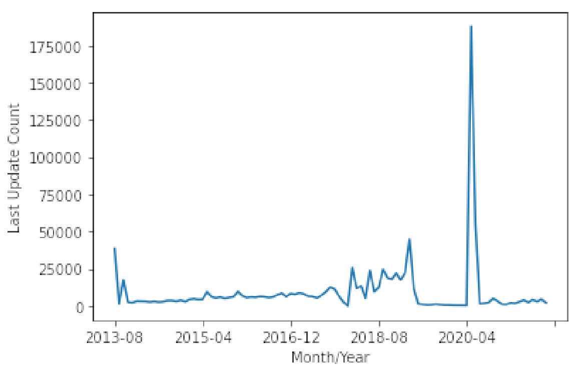

Figure 3: Count of places globally with a last update within a particular month.

Our first idea was to find places that have been marked in a short interval from each other. This experiment failed, as we found out that the 'updated_at' field in the Wheelmap data—which is how we hoped we could tell when a location has been marked—is set by any update (to that place's record) in the system. This is depicted in Fig.3, where the peak update possibly corresponds with some internal Wheelmap database migration. In the next stage of our research, we plan to look into the historical OSM data to detect when locations have being marked.

Our second idea was to find blocks within regions that have a high percentage of marked places close to each other. Unfortunately this experiment did not succeed either. The limitation might be that not all types of place are marked during a mapathon. In other words, some type filtering might be necessary for detection. We plan to interview a few mapathon attendees to better understand what happens during a mapathon (and how it reflects in the data).

Filtering certain place types might be good for other reasons as well. Wheelmap, and the underlying OSM data, list over 150 different place types. A test of their distribution across regions (using the coefficient of variation metric and after correcting for size) reveals very high variance. Part of this variance can be explained by geography, e.g., not all regions have caves. But it probably also reflects different mapping conventions. Thus filtering, as well as re-categorizing some place types, might be necessary to correctly estimate the links between `Place Attributes` and other nodes.

A final idea is to find ground truth on accessibility from other sources, and estimate the link between `isAccessible` and `isMarked` using that. We have not yet found such a dataset and would appreciate any pointers and hints from the community (even if the data is limited to only some regions). Such data would also aid us during cross-validation.

4.3 Evaluating policy interventions

To give a taste of what is possible once we have a proper approximation of access, we present a regression model that estimates the number of accessible places in a region using a number of factors (Table 2). The regional variables, selected from the latest EuroStat[8] data, along with a basic rational for their inclusion, are as follows:

[8] https://ec.europa.eu/eurostat/estat-navtree-portlet-prod/BulkDownloadListing

Table 2: Negative-Binomial Regression Results
Dep. var: accessible points (in a region)

Independent variable	Coef.	Std. err.	Z	P>|z|
urban	0.193	0.12	-1.60	0.11
poverty	-0.042	0.01	-3.81	0.00
tourism	0.157	0.07	2.32	0.02
spi	0.071	0.010	7.03	0.00

Dep.var. offset by population; N:237; Pseudo RSq:0.03; Alpha:0.93

- *Urban*: whether a region is predominantly urban (1), rural (-1), or mixed (between -1 and 1); our hypothesis is that places in urban areas should typically be easier to make accessible.
- *Poverty*: percentage of people at risk of poverty or social exclusion; our hypothesis is that poorer region will have less resources for modernizing and increasing access.
- *Tourism*: nights spent at tourist accommodation establishments in a region (logged); our hypothesis is that more popular destinations are more likely to be mapped and accessible.
- *Social Progress Index (SPI)*: measures the extent to which countries provide for the social and environmental needs of their citizens (the index is between 0 and 100 and based on 55 underlying indicators[9]); our hypothesis is that regions with a higher SPI should, on average, have higher accessibility.

Such regression results are what one would see in more a classic econometric paper.

Caution is necessary when interpreting the regression results. Foremost, the dependant variable is a biased lower-bound estimate of the true number of accessible places, as should be hopefully clear by this point. Secondly, the regional variables need a DAG themselves, before any causal claims can be made, due to possible confounders. Lastly, it may be more useful to approximate accessibility of regions at a granularity higher than NUTS2 regions—maybe even at the level of city blocks.[10]

In future work, it will also be interesting to look at the effectiveness of policies intending to increase access, in addition to structural socioeconomic and geographical factors.

5 RELATED WORK

Accessibility and assistive technology for PWDs have been a growing area of research over the last decades, in particular for blind or low vision people, people with motor/physical disabilities and people who are deaf or hard of hearing.[7] There has also been a considerable amount of research aimed at automating the documentation of accessibility in the built environment; yet so far, no fully automatic system that has been shown to reliably document surface quality barriers in the built environment in real-time.[6]

Crowdsourced and community maps have been successful tools, especially during humanitarian response, such as after the Haiti and Nepal earthquakes.[3] Through an analysis of 51 mapping deployments between 2010 and 2016, [3] point out that organisational structure matters for effective mapping actions, and suggest regionalising, preparation, and more research for further improvements. [5] looks at the behavior of the mappers who contribute to OSM (using both network analysis techniques and qualitative analysis), and identify a number of distinct mapping practices. Work that aims to understand the bias introduced by crowdsourcing mapping to volunteers include [13], which quantifies content bias across a three-year period of OSM mapping in 40 countries; it concludes no content bias exists in terms of what has been mapped, but finds significant bias in geography and meticulousness. In our opinion, these works add weight to the idea of filtering and normalizing the underlying OSM data before analysis of the additional Wheelmap tags.

The promise and impact of citizen generated open data in public governance is discussed in [8], with a study of 25 cases in different countries; they argue that the contribution of citizen-generated data to public governance should be understood in terms of both collaboration and contestation. Our own take is that many of these other projects (which generate open data with the help of citizens) will similarly deal with the problem of selection bias. Consequently, the causal approach we propose in this work might be useful for these other projects as well.

The causal inference framework presented in Section 3 was developed within the AI community [10]. There has been an increasing interest in recent years in its application in other fields, such as epidemiology[12] and economics[4]. The latter work offers a historical overview of the developments of the causality literature. To the best of our knowledge, our paper is one of the first attempts of applying the causal inference framework within HCI and accessibility research.

6 CONCLUSION

Wheelmap.org has had tremendous success in mapping accessibility at scale, soliciting information on the accessibility of over a million places since its start. Through the course of this paper, we discussed several use cases for approximating the accessibility of areas and regions based on this data, with the aim of ultimately empowering PWDs and aiding progressive policies. However, as is the case with any metric and score, one must be careful not to confuse what is quantifiable with what is real.

And here in lies the key challenge: despite Wheelmap's impressive breadth, it still covers only 10% of all public places within Europe—the continent with the best coverage. Using the framework of causal inference, we argued that the selection bias resulting from the volunteers mapping choices lead to unreliable approximations—a point that is also reflected in the empirical data. Luckily, the causal framework also offers tools to adjust for the missing data, for instance, by using information about mapathons or collecting some ground truth, the approximations can be improved.

We look forward to receiving feedback from the community as we take further steps in this research journey.

ACKNOWLEDGMENTS

The authors wish to thank Wheelmap.org for kindly providing us with the data used in this research. This research has been funded by the German Ministry for Education and Research (BMBF).

[9]https://ec.europa.eu/regional_policy/en/information/maps/social_progress2020/
[10]This makes sense if one considers that the business and residential characteristics of neighborhoods differ, and so does the age of the infrastructure within them.

REFERENCES

[1] Elias Bareinboim and Judea Pearl. 2016. Causal inference and the data-fusion problem. *Proceedings of the National Academy of Sciences* 113, 27 (July 2016), 7345–7352. https://doi.org/10.1073/pnas.1510507113 Publisher: National Academy of Sciences Section: Colloquium Paper.

[2] Judith Heumann. 2016. Our fight for disability rights – and why we're not done yet. https://www.ted.com/talks/judith_heumann_our_fight_for_disability_rights_and_why_we_re_not_done_yet

[3] Amelia Hunt and Doug Specht. 2019. Crowdsourced mapping in crisis zones: collaboration, organisation and impact. *Journal of International Humanitarian Action* 4, 1 (Dec. 2019), 1. https://doi.org/10.1186/s41018-018-0048-1

[4] Paul Hünermund and Elias Bareinboim. 2021. Causal Inference and Data Fusion in Econometrics. arXiv:1912.09104 [econ.EM]

[5] Marina Kogan, Jennings Anderson, Leysia Palen, Kenneth M. Anderson, and Robert Soden. 2016. Finding the Way to OSM Mapping Practices: Bounding Large Crisis Datasets for Qualitative Investigation. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. Association for Computing Machinery, New York, NY, USA, 2783–2795. https://doi.org/10.1145/2858036.2858371

[6] Marvin Lange, Reuben Kirkham, and Benjamin Tannert. 2021. Strategically Using Applied Machine Learning for Accessibility Documentation in the Built Environment. In *Human-Computer Interaction – INTERACT 2021*. Springer International Publishing, Cham, 426–448.

[7] Kelly Mack, Emma McDonnell, Dhruv Jain, Lucy Lu Wang, Jon E. Froehlich, and Leah Findlater. 2021. What Do We Mean by "Accessibility Research"?: A Literature Survey of Accessibility Papers in CHI and ASSETS from 1994 to 2019. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–18. https://doi.org/10.1145/3411764.3445412

[8] Albert Meijer and Suzanne Potjer. 2018. Citizen-generated open data: An explorative analysis of 25 cases. *Government Information Quarterly* 35, 4 (2018), 613–621. https://doi.org/10.1016/j.giq.2018.10.004

[9] OpenStreetMap. nd. Wheelmap - OpenStreetMap Wiki. https://wiki.openstreetmap.org/wiki/Wheelmap

[10] Judea Pearl, Madelyn Glymour, and Nicholas P. Jewell. 2016. *Causal Inference in Statistics - A Primer* (1st edition ed.). Wiley, Chichester, West Sussex.

[11] Judea Pearl and Dana Mackenzie. 2018. *The Book of Why: The New Science of Cause and Effect*. Hachette Book Group, New York, NY, USA. https://www.amazon.com/Book-Why-Science-Cause-Effect/dp/1541698967/

[12] Maya L. Petersen and Mark J. van der Laan. 2014. Causal models and learning from data: integrating causal modeling and statistical estimation. *Epidemiology (Cambridge, Mass.)* 25, 3 (May 2014), 418–426. https://doi.org/10.1097/EDE.0000000000000078

[13] Giovanni Quattrone, Licia Capra, and Pasquale De Meo. 2015. There's No Such Thing as the Perfect Map: Quantifying Bias in Spatial Crowd-sourcing Datasets. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '15)*. Association for Computing Machinery, New York, NY, USA, 1021–1032. https://doi.org/10.1145/2675133.2675235

[14] SOZIALHELDEN e.V. nd. Wheelmap.org. https://sozialhelden.de/en/wheelmap-org/

[15] TEDx Berlin. 2010. Raul Krauthausen- Wheelmap.org. https://www.youtube.com/watch?v=I13yL8ygwhk

[16] United Nations. 2006. Convention on the Rights of Persons with Disabilities and its Optional Protocol (A/RES/61/106). https://www.un.org/development/desa/disabilities/convention-on-the-rights-of-persons-with-disabilities/article-9-accessibility.html

[17] Wheelmap. 2021. Map My Hood - Wheelmap.org - New campaign. https://news.wheelmap.org/en/map-my-hood/