



The BALTO Toolkit - A New Approach to Ethical and Sustainable Data Collection for Equitable Public Transit

Saad Mohammad Abrar*

sabrar@umd.edu

University of Maryland, USA

Naman Awasthi*

nawasthi@umd.edu

University of Maryland, USA

Sunyup Park

sypark@umd.edu

University of Maryland, USA

Jessica Vitak

jvitak@umd.edu

University of Maryland, USA

Vanessa Frías-Martínez

vfrias@umd.edu

University of Maryland, USA

ABSTRACT

In most American cities commuters on public transit have disproportionately lower incomes than commuters who use automobiles. Given the proven link between geographic and economic mobility, it is critical to offer quality public transit to improve access to jobs, health care and education opportunities. Departments of Transportation (DOTs) routinely measure public transit performance and quality perceptions to assess the need for improvements in the transit systems. Nevertheless, the performance metrics used fail to capture the experiences of low-income individuals who often endure complex, lengthy trips, requiring several modes or transfers. We propose BALTO, a novel toolkit to characterize transit system performance and passenger's quality perceptions across all types of passengers and trips. We are designing the BALTO toolkit in collaboration with public housing residents from the Housing Authority of Baltimore City (HABC) and together with two local transit advocacy groups and the departments of transportation in Baltimore and in the state of Maryland.

CCS CONCEPTS

• **Human-centered computing** → **Mobile devices; Ubiquitous and mobile devices**; • **Information systems** → **Location based services**.

KEYWORDS

public transit, data collection, equity, sustainability

ACM Reference Format:

Saad Mohammad Abrar, Naman Awasthi, Sunyup Park, Jessica Vitak, and Vanessa Frías-Martínez. 2023. The BALTO Toolkit - A New Approach to Ethical and Sustainable Data Collection for Equitable Public Transit. In *ACM SIGCAS/SIGCHI Conference on Computing and Sustainable Societies (COMPASS '23)*, August 16–19, 2023, Cape Town, South Africa. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3588001.3609374>

*These authors contributed equally to this research.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

COMPASS '23, August 16–19, 2023, Cape Town, South Africa

© 2023 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-0149-8/23/08.

<https://doi.org/10.1145/3588001.3609374>

1 INTRODUCTION

Most American cities with substantial public transit ridership share a stark statistic: commuters on public transportation have disproportionately lower incomes than commuters who use automobiles [5, 10, 18]. Previous research has also shown that higher income residents who use public transit typically rely on single-boarding trips, while lower-income individuals endure complex, lengthy trips, requiring several modes or transfers [15]. Put together, these two facts reflect a socio-economic disparity in both public transit use and overall service quality. Underpinning this relationship is extensive literature demonstrating the link between geographic and economic mobility; those who can move around easily, have higher chances of improving their economic status via access to jobs, health care and education opportunities [3, 9].

Prior research with low-income communities in Baltimore (US) validated these findings [11, 12]. Residents expressed frustration in the perceived low-quality public transit service that limited access to work and educational opportunities, and a feeling that other, more affluent communities had better service. With approximately 15% of public transit commuters falling in the poverty bracket and with commuter average salaries 33% lower than those of non-public transit commuters ($\approx 20K$), Baltimore's transit system is one of the most class-segregated in the US [5, 17]. Although commuting time has emerged as the single strongest factor in the odds of escaping poverty in metropolitan areas - well above affordable housing, public safety or access to better schools [4] - some of the most equity-disadvantaged neighborhoods in Baltimore have the highest percent of people using public transit and traveling more than 45 minutes to work [8].

Departments of Transportation (DOTs) routinely measure public transit performance and quality perceptions to assess the need for improvements to the transit system. Nevertheless, we posit that the metrics currently used are not sufficient to understand the barriers that low-income residents face; which can in turn negatively impact the equity of the public transit. DOTs measure transit quality using two approaches: vehicle-centric measures whereby GPS data from onboard trackers (e.g., bus trackers) is used to measure on-time performance; and user-centric measures whereby passengers are invited to participate in quality of service surveys (QoS) that collect public transit performance perceptions from users of the transit system. Although both approaches are widely used, they suffer from two important limitations. First, vehicle-centric measures can characterize the performance of a specific route e.g., bus 71, but fail

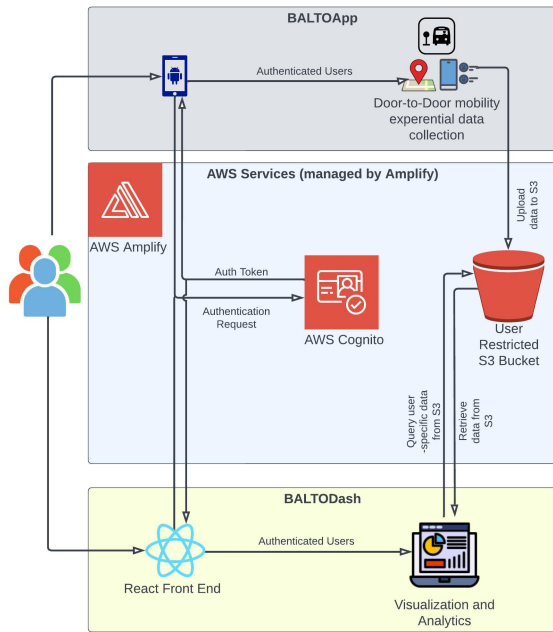


Figure 1: BALTO Toolkit Architecture.

to identify the performance issues that might arise during multi-modal trips, when passengers use more than one type of public transit *e.g.*, bus and metro; or during multi-legged trips *i.e.*, when passengers make one or more transfers on a given public transit type *e.g.*, transfer from one bus to another. By collecting vehicle performance data per public transit route, the information around multi-modal and multi-legged experiences is lost. For example, two buses might always be on-time but their connection infrequent, thus making trips that connect both routes unnecessarily long. This failure to collect multi-modal and multi-legged experiences results in lack of representation of trip experiences from those with more complex trips, who tend to be low-income residents.

Second, user-centric measures are computed via passenger surveys that are either carried out aboard per individual route *e.g.*, surveying passengers on a bus; or online after the trip, by encouraging passengers to fill out a survey online *e.g.*, "Rate Your Ride" in Baltimore City [16]. Nevertheless, these approaches - again - fail to collect perceptions for multi-modal and multi-legged trips that are most commonly taken by low-income passengers. In other words, current approaches to measuring performance and quality of service often times fail to collect the experiences of lower-income residents using public transit, thus keeping their experiences and voices away from policy decisions and public transit system changes, and potentially making public transit less equitable.

In this demo, we present the BALTO toolkit (Be An Advocate for Public Transportation). The toolkit propounds a novel approach to collecting relevant public transit data to characterize system performance and passenger's quality perceptions across all types of trips from simple to the more complex multi-modal or multi-legged trips that low-income residents tend to experience at higher rates.

We are designing the BALTO toolkit in collaboration with public housing residents from the Housing Authority of Baltimore City (HABC) and together with two local transit advocacy groups CMTA and BTEC, and the departments of transportation at Baltimore (BCDOT) and the state of Maryland (MDOT MTA). By focusing on collecting door-to-door individual mobility experiences rather than on the traditional per-route metrics, the information collected by Baltimoreans with the BALTO toolkit will provide a holistic picture of both performance and quality perceptions across population from all income levels in Baltimore City.

The BALTO toolkit is being designed and developed with two important values at its core: ethical data collection and sustainability. The ethical data collection focuses on understanding the processes that residents feel comfortable with when collecting their own door-to-door mobility experiences. Sustainability focuses on designing the toolkit in a way that minimizes costs to maximize adoption by organizations with diverse budgets *e.g.*, DOTs with large budgets versus transit advocacy groups or community associations; as well as minimizing end user costs *e.g.*, data or battery usage. Lowering the barrier to data collection will help many communities and organizations collect the data they need to advocate for equitable public transit in their cities and communities. Next, we describe the BALTO toolkit in Section 2, the ethical design in Section 3 and sustainability in Section 4.

2 THE BALTO TOOLKIT

The BALTO toolkit consists of two main components: BALTOApp and BALTODash. We have developed BALTOApp, and we are currently working on BALTODash. Fig 1 shows the toolkit architecture.

BALTOApp. BALTOApp is a smartphone app that allows residents to collect high-resolution, individual, door-to-door mobility experiences. These experiences consist of GPS data characterizing the complete trip from origin to destination *i.e.*, including all trip legs taken in public transit, as well as the first- and last-mile traveled to get to and from the public transit stop since, often times, public transit barriers appear in the form of accessibility to the public transit itself *e.g.*, having to walk a long distance to take a bus. Door-to-door mobility experiences collected by residents using the BALTOApp will also include QoS surveys - quality of service perceptions - that residents will be requested to fill out for each trip leg they have taken. This design will allow to unlock, for the first time, the full spectrum of complexities faced by transit users in multi-modal and/or multi-legged trips with multiple transfers.

BALTODash. BALTODash is designed as a dashboard that provides all stakeholders the ability to identify, characterize, and analyze neighborhood-centric problems with the public transit service. BALTODash will provide aggregate, neighborhood-level statistics and visualizations around the experiences of Baltimoreans while using public transit. Baltimoreans will be incentivized to collect their door-to-door mobility experiences on public transit for two weeks. After that data collection period, the dashboard will be used during facilitated, neighborhood focus groups with residents, advocacy groups and decision makers from departments of transportation - who are all partners in this project - to ignite conversations about transit challenges and potential solutions. Our focus groups research will assess the role that various conditions, processes, tools

and data play in the identification of data-driven, consensual solutions to transit challenges among stakeholders.

Cloud services. The BALTOApp and the BALTODash are integrated with AWS cloud functionalities using AWS Amplify, which provides a suitable development environment to use most of the AWS services. AWS Cognito is used to manage the identity pools and user management services. Individuals are authenticated using Amplify's Auth API, which helps connect the cognito user pools, hence handling the user authentication and authorization flows in the BALTOApp. Once individuals are authenticated, they can start collecting door-to-door mobility experiences. Amplify's Storage API helps upload the data collected to AWS S3 buckets (data lake), and manages the access modifiers with respect to the users in the cognito user pool ensuring the secure transfer of data. When users access BALTODash, the dashboard will request user-specific data from the corresponding S3 buckets to display insights and data-driven visualizations of both individual trips as well as neighborhood-centric aggregate experiences.

3 ETHICAL DATA COLLECTION

Research has shown that low-income groups tend to be more concerned than their higher income counterparts about how their personal information is collected and used [2], and less likely to use privacy settings when they navigate online [13]. However, more research is needed to understand their privacy perceptions and attitudes with respect to location data collection on mobile apps; and with respect to the barriers posed by the combination of mobile surveys - that collect personal information - with location data, as done in the BALTO toolkit. We know that privacy is understood differently across social groups or technologies, making findings from other contexts or applications not generalizable to new settings [14, 19]. Our focus on transit-dependent, equity-challenged communities, and the complexities of the toolkit - the collection of location and QoS surveys in BALTOApp and the information flows between stakeholders in BALTODash - call for an in-depth inquiry of privacy perceptions in our setting so as to guide the design of both BALTOApp and BALTODash. Specifically, we are interested in (1) understanding the privacy barriers posed by the collection of mobility experiences and by data sharing processes across stakeholders and tools; and (2) the design guidelines needed to lower those barriers. We are answering these questions by running focus groups with low-income residents to understand their privacy concerns in the context of the BALTO toolkit; and based on findings from the focus groups we are proposing design guidelines to address the concerns identified. We have started this line of inquiry with a focus on the BALTOApp. Next we describe some of our preliminary findings and how we use them to guide the design of ethical data collection in the BALTOApp.

3.1 Understanding Privacy Barriers

We have carried out 10 focus groups with a total of 45 participants from three public housing locations in Baltimore, in collaboration with our project partner, the Housing Authority of Baltimore City (HABC). Figure 2 shows the three locations in a map: Bernard E. Mason Apartments, Cherry Hill Homes and Rosemont. Next, we summarize our main findings:

- Participants were generally less concerned with their location data being collected, as long as they understand that it is necessary for the transit app to function.
- Nevertheless, participants wanted minimum data collection, and a clear and detailed explanation as to why their location data is being collected.
- Additionally, participants wanted to be able to manage the data being collected, such as having the ability to control their location data according to granularity.

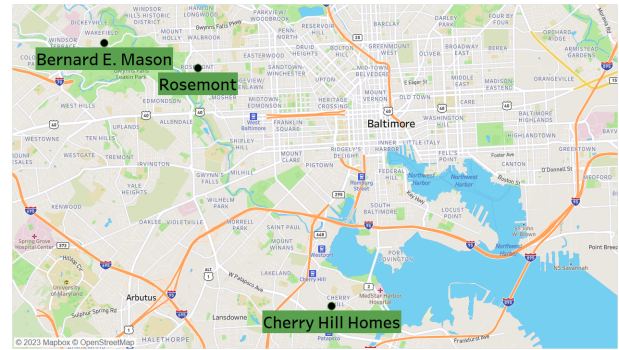


Figure 2: Public Housing Locations for Focus Groups (HABC).

3.2 Design Guidelines

We have used the feedback distilled from the focus groups to assist in the design of ethical data collection for the BALTOApp. Specifically, we have incorporated trip controls that will give residents control over their privacy and the data they would like to share with the BALTOApp; and we have devised a protocol for the use of the toolkit components (app and dashboard). Next, we describe each design guideline in detail:

- Incorporate a stop button in the BALTOApp. This button will allow residents to stop the collection of their GPS data for a given period of time during their trip. This allows residents to prevent data collection during *sensitive, private* parts of their trip that they do not want to share e.g., visit to a doctor.
- Incorporate a cancel button in the BALTOApp. This button will allow residents to stop and delete the collection of GPS data for a given trip. Residents will be able to decide, at any time during the trip, whether they would like to share a given door-to-door experience with the BALTOApp. Internally, BALTOApp does not upload GPS data to the cloud until the end of the trip; upon use of the cancel button, the GPS data saved locally in the phone is deleted and no data is uploaded to the cloud.
- Deploy BALTOApp and BALTODash services concurrently so that residents can collect their door-to-door mobility experiences with BALTOApp and then use BALTODash to transparently explore their own mobility experiences, and to understand how their trip information is aggregated to the neighborhood level to be explored during the neighborhood focus groups with residents, advocacy groups and decision makers.



Figure 3: (a) Enter Trip Information; (b) Trip Controls; (c) Surveys shown in the App.

3.3 BALTOApp UI

The BALTOApp has been designed and implemented according to the data collection needs described in the Introduction, and following the design guidelines described in the previous section. Figure 3 shows a BALTOApp prototype that includes three main components: (1) the screen to input trip details, (2) the screen with the trip controls, and (3) the screen to fill out a QoS survey. This prototype has been implemented on Android, and is currently available for download in the Google App Store [20]. We have run usability tests for this prototype with six residents from one HABC location (Cherry Hill Homes) and the prototype was viewed as a good channel to record experiences on public transit. We will continue our design and implementation cycle by incorporating future findings from additional focus groups and usability tests into the BALTOApp, and by designing and implementing the BALTODash with community members.

4 TOOLKIT SUSTAINABILITY

We have made several decisions in the design of the toolkit to maximize sustainability by minimizing different types of costs across users including public transit passengers, advocacy groups and transit decision makers. Next, we describe each decision in detail.

AWS Costs. The design of the BALTO toolkit is centered around minimizing AWS costs so that advocacy groups, who often times are non-profit organizations with limited funding, can also use the toolkit. Transit decision makers in departments of transportation might have larger budgets available, but our design is focused on inclusivity across all organizations who can have a voice in transit decisions. Thus, we have made several strategic choices for AWS:

- We are using an AWS S3 Standard bucket as our data lake to upload the GPS data and QoS survey data collected by residents. Instead of paying for a dedicated infrastructure, we only pay for the storage and data transfer for each trip uploaded to S3. Assuming 1,000 resident participants with an average of two 1-hour trips per day, the total cost of our approach is \$5/year.
- BALTODash will be a locally hosted webapp that displays public aggregate transit experiences at the neighborhood level. To

accomplish that, we will design novel machine learning (ML) methods to identify shared patterns across the trips collected, such as long waits at a transfer, long first- or last-mile trips, or complaints related to the cleanliness of the transit vehicles. Training ML methods and running inferences on AWS is expensive. We considered using dedicated ML platforms - like SageMaker - or deploying our own ML methods on an EC2 server. Nevertheless, the costs to process data from 1,000 participants can go up to, at least, \$30/month. Thus, we have decided to implement a zero-cost automatic pipeline that will (i) download the data from S3 to run the ML methods on a local machine, and then (ii) upload the data to S3 again to make it available for BALTODash.

Battery Costs. To maximize participation in the collection of door-to-door mobility experiences among public transit users, BALTOApp has been designed with a focus on minimizing battery consumption given that battery drainage is an important concern in smartphone use [6]. For that purpose, we have made the following design decisions.

- The GPS data and QoS surveys are uploaded to S3 at the end of the trip, instead of uploading the information in real time, which can help reduce battery consumption [1].
- GPS data is collected using the Google Play Service Fused Location Provider API. This service is optimized to balance battery utilization and accuracy of location [7] by extrapolating the best location estimations using GPS, Wi-Fi and the mobile network.
- QoS surveys are pushed at the end of each trip leg using geofences. BALTOApp adheres to best battery-efficient practices in the Android geofence API developer notes initiating geofences only when a stop is reached and using large radii ($300m \leq r \leq 5km$).

Overall, BALTOApp increases battery consumption by 1%-3% (absolute increase) for a one-hour trip when sampling locations at a rate of 0.008Hz (1 per 120 seconds). This empirical result was derived from results of three devices (Google Pixel, Google Nexus and OnePlus) with different OS specifications.

5 CONCLUSION AND FUTURE WORK

Access to reliable and efficient public transit is one of the most significant needs in Baltimore City and in metropolitan areas across the US. We posit that the BALTO toolkit will allow low-income residents, advocacy groups, and decision makers to collect the information necessary to identify high value, small scale, community-led improvements for the current transit network. These improvements will, in turn, offer significant chances for augmented access to economic opportunity and growth, health, and overall quality of life, especially for those who are currently most disadvantaged by the transit system. Future work will focus on the design and development of the BALTODash, on minor improvements on the BALTOApp, and on deploying the toolkit together with community associations in Baltimore.

REFERENCES

- [1] Marco V Barbera, Sokol Kosta, Alessandro Mei, and Julinda Stefa. 2013. To offload or not to offload? the bandwidth and energy costs of mobile cloud computing. In *2013 Proceedings IEEE Infocom*. IEEE, 1285–1293.
- [2] Pew Research Center. 2019. Americans and Privacy: Concerned, Confused and Feeling Lack of Control over their Personal Information. <https://www.pewresearch.org/internet/2019/11/15/americans-attitudes-and-experiences-with-privacy-policies-and-laws/>. accessed April 2023.

- [3] Nathan Iel; Kline Patrick; Chetty, Raj; Hendren and Emmanuel Saez. 2015. Economic Mobility. Pathways: The Poverty and Inequality Report. *The Stanford Center on Poverty and Inequality* (2015).
- [4] Raj Chetty and Nathaniel Hendren. 2018. The impacts of neighborhoods on intergenerational mobility II: County-level estimates. *The Quarterly Journal of Economics* 133, 3 (2018), 1163–1228.
- [5] Citizens Planning and Housing Association, Inc. 2014. Who uses public transportation in Baltimore? <http://www.cphabaltimore.org/2014/04/who-uses-public-transportation-in-baltimore/>. accessed May 2022.
- [6] Denzil Ferreira, Anind K Dey, and Vassilis Kostakos. 2011. Understanding human-smartphone concerns: a study of battery life. In *Pervasive Computing: 9th International Conference, Pervasive 2011, San Francisco, USA, June 12-15, 2011. Proceedings* 9. Springer, 19–33.
- [7] Google Developers. [n. d.]. Fused Location Provider API. <https://developers.google.com/location-context/fused-location-provider>. accessed May 2023.
- [8] Seema Iyer. 2018. Lack of accessibility leads to high-commute time neighborhoods (Research contribution - opinion piece invited by cmta). <https://medium.com/central-maryland-transportation-alliance/lack-of-accessibility-leads-to-high-commute-time-neighborhoods-28028c0f1197>. accessed April 2023.
- [9] Sarah Kaufman, Mitchell L Moss, Justin Tyndall, and Jorge Hernandez. 2014. Mobility, economic opportunity and New York City neighborhoods. *NYU Wagner Research Paper* 2598566 (2014).
- [10] Michael Kodransky and Gabriel Lewenstein. 2014. Connecting Low-income people to opportunity with shared mobility. https://itdpdotorg.wpengine.com/wp-content/uploads/2014/10/Shared-Mobility_Full-Report.pdf. accessed April 2023.
- [11] Willow Lung-Amam, Ariel Bierbaum, Sheri Parks, Lauren Stamm, Gail Sunderman, and Gerrit Knaap. 2018. Smart Cities, Connected Communities: Using Technology to Meet the Needs of the West Baltimore Residents. <http://www.umdsmartgrowth.org/wp-content/uploads/2018/11/Final-Smart-Cities-West-Baltimore-Engagement-Team-Report.Nov-2018.pdf>. accessed April 2023.
- [12] Willow Lung-Amam, Ariel H Bierbaum, Sheri Parks, Gerrit-Jan Knaap, Gail Sunderman, and Lauren Stamm. 2021. Toward engaged, equitable, and smart communities: Lessons from west Baltimore. *Housing Policy Debate* 31, 1 (2021), 93–111.
- [13] Mary Madden, Lee Rainie, Kathryn Zickuhr, Maeve Duggan, and Aaron Smith. 2014. Public perceptions of privacy and security in the post-Snowden era. *Pew Research Center* 12 (2014).
- [14] Andrew R. McNeill, Lynne Coventry, Jake Pywell, and Pam Briggs. 2017. Privacy Considerations when Designing Social Network Systems to Support Successful Ageing. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) (*CHI '17*). ACM, New York, NY, USA, 6425–6437. <https://doi.org/10.1145/3025453.3025861>
- [15] Cornelius Nuworsoo, Elizabeth Deakin, and Aaron Golub. 2012. Equity Impacts of Transit Fare Proposals: A Case Study of AC Transit. In *Journal of the Transportation Research Forum*, Vol. 47.
- [16] Maryland Department of Transportation (MDOT). [n. d.]. Rate Your Ride. <https://rateyourride.org/>. accessed April 2023.
- [17] Pew Research Center. 2016. Who relies on public transit in the US? <https://www.pewresearch.org/fact-tank/2016/04/07/who-relies-on-public-transit-in-the-u-s/>. accessed April 2023.
- [18] Populus. 2018. The Micro-mobility Revolution. <https://www.populus.ai/micro-mobility-2018-july>. accessed April 2023.
- [19] Yang Qin, Bin Xu, and Dan Cosley. 2017. Designing the Interplay between Anonymity and Publicity for Online Social Support. In *Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. ACM, 283–286.
- [20] UMD. [n. d.]. BALTO App. <https://play.google.com/store/apps/details?id=balto.loc.kotlin>. accessed July 2023.