Developing an Interactive Landscape for Mobility Resources

Preprint

Lauren Janicke,¹ Benjamin Burch,² K. Shankari,³ and Arthur Yip³

¹ Carnegie Mellon University
² Temple University
³ National Renewable Energy Laboratory

Presented at the Transportation Research Board (TRB) 101st Annual Meeting
Washington, D.C.
January 9-13, 2022
Developing an Interactive Landscape for Mobility Resources

Preprint

Lauren Janicke,1 Benjamin Burch,2 K. Shankari,3 and Arthur Yip3

1 Carnegie Mellon University
2 Temple University
3 National Renewable Energy Laboratory

Suggested Citation
NOTICE

This work was authored in part by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. This work was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internship (SULI) program. This work was also supported in part by the Laboratory Directed Research and Development (LDRD) Program at NREL and the U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy, Vehicle Technologies Office. The views expressed herein do not necessarily represent the views of the DOE or the U.S. Government. The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. Government purposes.

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.


Cover Photos by Dennis Schroeder: (clockwise, left to right) NREL 51934, NREL 45897, NREL 42160, NREL 45891, NREL 48097, NREL 46526.

NREL prints on paper that contains recycled content.
ABSTRACT
As the world continues to be increasingly driven by data, the ways researchers and professionals sort and collect these data are critical. In the world of mobility data, new levels of data from public transportation systems, location services, and other means are being lost due to how little organization exists. Much of the data are proprietary, and there are few if any de jure or even de facto standards connecting data. There is also little knowledge about the gaps that exist in the data. In this project, we created an interactive landscape where mobility resources are categorized and organized in an easy to use, living document. We made this landscape with open-source code from the CNCF Cloud Native Landscape and repurposed it to the mobility data’s needs. Additionally, unlike previous sources that organize mobility data, this document can be updated through GitHub by those in the field to keep its sources relevant. Following the creation of a beta version of the landscape, we conducted several interviews with industry researchers and professionals to ensure the landscape would be useful. The result is an online hub where mobility researchers and resource creators can easily access research and collaborate.

Keywords: Data, Standards, Living Landscape, Mobility
INTRODUCTION

Data science is becoming one of the most impactful research methods, regardless of the field. The ability to use large quantities of new data to survey topics has brought compelling research to older fields. Mobility, the movement of people or vehicles, is one of many fields advanced by data science. With the explosion of location broker services, micro-mobility, and the recording of modernized public transportation systems, more mobility data exists now than ever before. However, while the data exist to be used for mobility data analysis, it is often difficult for researchers to find out what data are available and where to find them. Various organizations, from the government to private companies, collect these data, and there are little to no consensus on specifications for these data. We created a novel online hub, the Interactive Mobility Resource Landscape (https://nrel.github.io/mobility_landscape/), where researchers and resource creators can access and collaborate on mobility data and other resources.

The Interactive Mobility Resource Landscape is a living online document, based on the open-source code from the Cloud Native Computing Foundation Landscape (1), a similar project made to organize cloud-based computer programs. Any researcher or data collector can submit their work to the landscape. This interactive feature will be key to the potential success of the landscape, as available data continues to grow. An example of this growth is current ridesharing data with apps such as Uber or Lyft. Previously their data were kept away from public use, but they have been required by local governments to release this data. With this landscape, as more mobility data are released, it will become increasingly easy to discover comparable data and identify gaps in knowledge (2).

The process for creating this novel landscape involved a major meta-data analysis of the available data, specifications, and tools existing in the mobility field. Broadly, we noticed that trends and gaps in the data regarding geographic range, consistency in reporting, and privacy concerns. For example, despite there being little to no de jure specifications across the mobility landscape, several de facto standards do exist. The most prominent of these examples include General Transit Feed Specification, or GTFS, which used compatibility with Google Maps as a bargaining chip to create a de facto standard that is used by most public transportation systems (3). As no previous studies or projects have attempted to categorize this wide-reaching field, our main goal at the National Renewable Energy Laboratory (NREL) was to create an organized structure that would be easy for users to understand and use.

RELATED WORK

As mentioned earlier, this mobility landscape is based on the CNCF landscape for cloud computing (1). The Linux Foundation AI & Data Foundation Interactive Landscape and the Urban Computing Foundation Interactive Landscape are also based on the CNCF landscape (4–5). These landscapes do not attempt to change the filtering system of the original CNCF landscape; however, we customized the original landscape code to better fit our needs.

Previous efforts to quantify the scope of mobility data were done in static lists, such as NREL’s report as a part of the Smart City challenge (6). Other active data repositories like Livewire, the Mobility Data Marketplace, and the Transportation Secure Data Center do not include mobility resources beyond data and do not attempt to be a comprehensive list of data sources (7–9). To the best of our knowledge, we are the only project that is a living document meant to share mobility data and resources.

METHODS

The first step to create the mobility landscape required editing the original CNCF landscape to fit our purposes. The filtering system was refined to capture various important aspects of mobility resources and data. We considered how to address data privacy issues in the landscape. We collected an initial set of entries and tested the first organized structure of categories and subcategories. These categories were an introduction to what was available in the field and sorted the original 40 entries. After this initial gathering of data, the next steps were to review the organization structure and identify gaps that might exist in the initial entries by talking to experts in the field. The Open Mobility Data Landscape Advisory Board was created from professionals in the mobility field, both to assess and add to the initial data collected as well as supervise the creation of the landscape to make sure that it would be of use to
researchers. The Open Mobility Data Advisory Board met biweekly to comment on and suggest improvements for the landscape. We incorporated feedback from the advisory board to reorganize the categories and subcategories of the landscape to be more specific and intuitive for users. Then, individual interviews were conducted with other mobility researchers to collect more feedback.

**Creating the Mobility Landscape**
The original landscape (Figure 1) was our inspiration and starting point for the landscape code. The code for the landscape is split between two repositories, titled mobility_landscape and mobility_landscapeapp. Both stored on GitHub, these two repositories interact to create the eventual landscape. This landscape is currently hosted on GitHub pages and available for anyone to view. The basic data, or access point, for the data is stored in a YAML file, where aspects such as the title, basic description, website, and other variables about the entry are stored. Researchers and specification creators can add to the landscape by pulling the YAML file and submitting changes for approval by whoever is supervising the landscape. Anyone can pull this file, update it with relevant information, and then submit it to the landscape so that it evolves with the mobility field. The landscape runs off these basic tools; however, changes to the Categories, Subcategories, Fields, and Filtration System are more complex.

![Figure 1 Original CNCF Landscape](image)

**The First Categories and Subcategories**
The first version of the landscape had three categories: Infrastructure, Vehicles, People. Infrastructure involved all aspects of mobility data that include nonmoving objects, including highways, bike routes, preplanned public transit routes, etc. Vehicles involved all data that showed how vehicles move, including public transportation data, highway data, bike sharing apps, etc. The People category involved all data that related to how people specifically move around, including location broker data, tools like Walkscore, etc. These categories included subcategories to further organize the data. The first subcategories were made to specifically fit the needs of the specific entry. For example, if the Federal Transit Administration (FTA) posted data on ridership, that data would fall into the People category, and then a subcategory named Ridership was created. However, as the number of entries began to grow, and the number of...
unique subcategories also grew, this plan became unwieldy for users as too many fields were populated with only a single entry.

In the next iteration of the first landscape’s categories, a compromise was made between a standardized format and the more custom version. All overlapping categories were combined into Infrastructure-Vehicles-People, as a measure to ensure that every category had enough entries within it. This change meant that there were four categories: Infrastructure, Vehicles, People, and Infrastructure-Vehicles-People. To address the issue of specificity within the subcategories, three standard subcategories were added: Data, Tools, and Apps. However, if there were at least three entries that suggested a unique subcategory, it was added (Figure 2). An example of this is the Crowdsourced Maps subcategory of Infrastructure with four entries. Additionally, if any of the subcategories didn’t have enough entries the subcategory was deleted.

Figure 2 The Organization of Version 1.2 of the Landscape

Filtering System
Following the foundational work with the entries, categories, and subcategories, the next step was to revise the filtration system. The filtration system is what allows a researcher to toggle between the different aspects of the entries (Figure 3). For example, if a researcher only wanted to use resources from Europe, they would select this in the filtration system, and the landscape would display only entries that meet that criterion (Figure 4). This filtering is achieved using the different variables collected in the
original landscape YAML file. Originally, the focus of the project was on the previously mentioned categories and subcategories for organization, but input from the Open Mobility Landscape Advisory Board showed a specific interest in having a thoroughly developed filtration system for the landscape. After several drafts and comments by the board, the final filtration system involves nine major filters, each with major and minor filters within. A user can select for multiple options at once, which means that someone using the system can select exactly what they are looking for within the landscape.

![Figure 3 The Filter and Subfilter Options](image)
Data privacy is one of the biggest limiting factors around mobility data and is an important aspect of the data for researchers. Privacy impacts almost every field of modern mobility data, particularly data about vehicles and people. Location brokers have caught media attention and raised concerns about sharing, consent, and control of data (10). Thus, privacy concerns have limited mobility data in two major ways.

First, companies use privacy concerns as a major reason to not release mobility data. This withholding is especially prevalent in the field of micro mobility like bike or scooter sharing and in the world of ride-sharing apps like Uber and Lyft; however, these actions are understandable given state privacy laws and profits to be made from licensing data. Recently, governments and researchers have begun to ask these companies to release their data for open use to promote more mobility research (11).

The second major limitation to mobility data is the standard or protocol for anonymizing the data. The release of data must be taken with care because researchers have shown that a shockingly small amount of data can identify particular people (12). Thus, the privacy measures taken are a critical aspect of modern mobility data. Some of the differing methods include unlinking the trajectory of the data, putting it into buckets to combine an area of data, and making the data “fuzzy” at a granular level to protect individuals. As these methods become more popular, additional ways to assess the privacy of data are emerging (13). We include privacy measures as a filter in our landscape to allow users to find data with appropriate anonymization.

The Second Categories and Subcategories
To address more of the Open Mobility Landscape Advisory Board’s concerns regarding the intuitiveness of the landscape, the landscape’s organization was revisited. The main issue is that the categories and subcategories were too broad. The broad categories created unwieldy visuals that would only be exacerbated as the landscape grows. Meanwhile, the broad subcategories failed to capture fundamental differences between resources. Not recognizing these differences also makes it more difficult to identify gaps in the landscape. The first draft of the landscape had the categories focus on the entry’s context in mobility, and the subcategories divide between resource types. For example, in the first landscape
(Figure 2), Infrastructure/Data contains both datasets and data specifications, and Infrastructure/Tools contains a database, Python libraries, and simulations. This organization reduced the specificity of each item. Now in the new landscape (Figure 5), the categories represent resource types, and the subcategories focus on context in mobility or further refine the resource type. Also, the vertical organization of the landscape represents the lifecycle of data: the collection, the standardization, the utilization. We believe this organization is more intuitive to users who are likely most concerned with the resource type.

Figure 5 The Organization of the Second Version of the Landscape

To make the categories easier for users to understand, we formally defined what an appropriate entry is for the landscape, the new categories, and the new subcategories. An appropriate entry directly captures, allows, or relates to the movement of people or passenger vehicles. As more entries are put in place by outside researchers, a system will have to be developed for the recommendation of new subcategories as well as just adding entries. There are currently seven categories: Data Collection
Methods, Data Sets, Data Repositories, Data Formatting, Tools, Agent-based Simulations, and Aggregate Simulations.

Data Collection Methods are ways to obtain data sets; the subcategories are Sensing, Sensing and Surveying, and Surveying. These subcategories are some of the most obvious of the landscape and do not prompt much confusion for organization. In the future, these subcategories may be refined to what type of sensors or surveys are used.

Data Sets are from a single source and contain information about the movement of people or passenger vehicles. Infrastructure may include data sets with information about when, where, and how many people move between non-moving objects. An example of an Infrastructure data set is Chargepoint because the company tracks when and which charging stations are being used (14). This information indicates how much electric vehicles are driven. People data sets may contain variables about when, where, how far, or how many people or households move. The National Household Travel Survey is a People data set because the survey has households as a common identifier when data are given about people, trips, and vehicles (15). The Vehicles subcategory includes data sets that capture when, where, how far, or how many vehicles move. Ride-sharing data are appropriate for this subcategory.

Data Repositories are a collection of multiple appropriate data sets. The Infrastructure, People, or Vehicles subcategories includes repositories with data sets that are largely related to the respective subcategory. The Transportation Secure Data Center is a People repository because most of the data sets are household surveys (Figure 6) (9).

Figure 6 Item View of the Transportation Secure Data Center

Data Formatting refers to ways to organize data sets and metadata. Specifications describe how data sets should be formatted to be compatible with each other. Standards describe the concepts behind the data’s content or management. Transmodel is a Standard because it defines public transportation
concepts and suggests data structures (16). A single data set’s metadata should not be in this category; it is meant for generalizations about multiple data sets.

Tools use or relate to mobility data to provide further context. Accessibility Metrics rate the ability of an agent to travel in an area. Crowdsourced Maps feature maps that contain information related to mobility data and are generally about infrastructure. Data Processors manipulate data typically for analysis or cleaning.

Agent-based Simulations are a large subset of Tools that create hypothetical results at the level of individual decision making or at a temporal scale less than or equal to daily. People describes simulations that have individuals or households as the agents. The Transportation Energy & Mobility Pathway Options tool falls under People Agent-based Simulation because it simulates household hourly decision making with data from the National Household Travel Survey even though the results are given at the county level (17). Routing provides the optimal directions from one place to another. Navigation contains entries that give real-time routing. Vehicles involve simulations that have vehicles as the agents.

Aggregate Simulations are a large subset of Tools that create hypothetical results at the level of population decision making or at a temporal scale greater than daily. Infrastructure accounts for simulations that estimate the usage of infrastructure due to mobility demands. Vehicle simulations account for how a vehicle fleet is used. The Automotive Deployment Options Projection Tool is an aggregate vehicle simulation because it does not simulate how much or where people drive; it uses annual miles traveled as an input to simulate vehicle purchases (18).

Interviews
The interviews were largely casual conversations to hear what the interviewees felt was important to discuss given their background. However, we did have some guiding questions that most interviewees were asked.

- If your project is in the Landscape, do you agree with where your project is located?
- How is your project currently integrated with other entries here?
- Do you see future possibilities for further integration for your project, or do you already have plans to become more integrated?
- Would you want to see visualizations of integration between projects and data sources?

We interviewed a variety of NREL researchers who work with mobility data to hear diverse perspectives. We talked with Lauren Spath Luhring and Nick Muerdter (Livewire), Leidy Boyce (Transportation Secure Data Center), Christopher Hoehne (Transportation Energy and Mobility Pathways Options tool, Mobility Energy Productivity Metric), Eric Wood (Electric Vehicle Infrastructure Projection Tool), and Venu Garikapati (Transportation Secure Data Center, Mobility Energy Productivity Metric) (7, 9, 17, 19-20). Also, talking to these people helps us get more potential users familiar with the purpose and entries of the landscape. Getting people to actually use the landscape will be essential to its success; users need to feel like there is a sense of community and that their resources will be used by others.

RESULTS & DISCUSSION
As this is a novel project, we expected there to an exploration phase when creating the landscape. With the landscape successfully created, we achieved the first goal of the project: to create a living hub where mobility resources can be congregated and accessed. The next step is to identify more gaps and limitations of the current landscape. The most prominent issue is how few entries there currently are. With only 90 entries, there are undoubtedly more mobility resources available. In particular, the current entries are heavily biased towards North American and European mobility sources, despite data existing for other countries. The hope is that this limitation will be filled by those working in the field as the landscape is released. It would be nearly impossible for a single research team to categorize every single resource with regards to the mobility field across the world.
Another success of the landscape is how it has exposed the current gaps within mobility research. The first draft of the landscape did not make the gaps between the availability of various resource types as apparent as the second draft. We made the gaps more evident by changing the organization of the landscape. Originally, the categories and subcategories were too broad to be able to easily identify the gaps. For example, in the Tools subcategory of Vehicles, there are simulations, data specifications, and a data scraper. With the more specific categories and subcategories of the new landscape, it is more obvious that few data repositories about the movement of people have been identified and that NREL heavily focuses on simulations rather than data collection for mobility (Figure 7). Another gap in the data came from a lack of information on how the data are reported. Variables including how often data are reported, what format the data are within, contacts for the data, related tools for the data, and more are all things that very few data reporters have knowledge of. What this means is that while the data can be found, the necessary information to collaborate with other researchers is often not reported. While the gaps regarding what resources exist are not as problematic as expected, some issues remain. Due to privacy concerns, there are still little data on personal mobility available. Also, there are very few de jure specifications established in the United States. With the landscape demonstrating these issues, we hope that those working in each respective field will be able to see the differences more accurately between their peers and collaborate to make their field more complete.

A survey sent out after the creation of the first version of the landscape revealed that most respondents saw the primary use of the landscape being for research in the mobility field, with publishing and collaboration equal as a secondary use. Other major highlights from the survey included the added suggestion of a search bar at the top, a hovering option when looking at entries, rearrangement of the subcategories, and improving display options when the filters are active.

Following the creation of the second landscape, all interviewees thought that the new organization was an improvement. People commented that it is easier to read through the rows rather than the columns. They believed that the landscape achieves our goal of trying to help people understand what resources exist already, how the resources are related, and the purpose of each resource. The domain knowledge of the interviewees also brought attention to some incorrectly organized entries. For example, Sigalert was in Data Collection Methods/Surveying; however, Boyce noticed that Sigalert focuses more on analyzing traffic reports than user-inputted data. In the future, we look to interview more people from different backgrounds like communications, private businesses, and others. These people will provide different insights to better develop the landscape as a pragmatic tool. Additionally, Garikapati, a manager of the Transportation Secure Data Center, mentioned that all the filters were likely not necessary. Given that there are 9 filters and 90 items, if you use all the filters, there will be very few entries left to consider. In the future, we plan to reevaluate the necessity of each filter and the utility they provide to the user.

The question of allowing duplicate entries entered the conversation during most interviews. One version of this question originated from interviewees’ agreement on the location of their project. All interviewees agreed with the placement of their project to a certain degree. Two people completely agreed. The other three felt like their project belonged in multiple categories, but our placement was still accurate. Interviewees also noted that some of the companies we included had products that fit multiple categories. This desire to allow duplication to some degree prompted a discussion of how to handle this issue. To address these issues, we decided to follow the lead of the CNCF landscape and allow companies with multiple products have an entry for each product; however, if a product fits multiple subcategories it should only go into one. A single project would likely fit under multiple subcategories only when it is a Data Set or a Data Repository. In these cases, we expect mobility researchers and resource creators to understand the possibility of finding an appropriate data set in another subcategory due to the ways data can be organized. For example, the National Household Travel Survey splits the data into files about trips, vehicles, and people; however, it is categorized as only a People data set because the household is a common identifier across all data sets (15).
Figure 7 The Current Interactive Mobility Resource Landscape (https://nrel.github.io/mobility_landscape/)
CONCLUSIONS

The successful creation of this project has resulted in a living landscape for mobility data moving forward. It can be updated by anyone in the field and has the potential to serve as a hub for research materials and other resources. The next steps for the landscape involve further improvements to the user experience and building its userbase. Including features like a search bar or hovering actions and refining the filtering system more will make the landscape more attractive to users. As more users are active in the landscape, more entries will be added which will further visualize gaps in the landscape. Success development of landscape will encourage collaboration between researchers and allow people to discover new resources.

ACKNOWLEDGMENTS

We would especially like to thank the Open Mobility Specifications Advisory Board for their suggestions on the project, our interviewees for their feedback, and the Mobility, Behavior and Advanced Powertrain group for supporting the research. This work was authored by the National Renewable Energy Laboratory (NREL), operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. This work was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internship (SULI) program. This work was also supported in part by the Laboratory Directed Research and Development (LDRD) Program at NREL and the U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy, Vehicle Technologies Office. The views expressed in the article do not necessarily represent the views of the DOE or the U.S. Government. The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. Government purposes.

AUTHOR CONTRIBUTIONS

The authors confirm contribution to the paper as follows: landscape creation idea and software changes: K. Shankari; first version of the landscape: B. Burch; reorganization idea: A. Yip; second version of the landscape: L. Janicke; draft manuscript preparation: L. Janicke, B. Burch. All authors reviewed the results and approved the final version of the manuscript.
REFERENCES


