

# 10 Principles of Mobility Data Operationalization

ENGAGEMENT MATERIAL  
JUNE 2020



# Contents

3	<b>1 Before you even get started</b>
4	Principle 1 Question your assumption
4	Principle 2 Take inventory, and take it often
5	Principle 3 There is no need to reinvent the wheel
6	Principle 4 Data likes (selective) company
7	<b>2 While you are at it</b>
8	Principle 5 Data is nothing but insights are everything
8	Principle 6 Less is most definitely not less
9	Principle 7 Ambition has many traps
11	<b>3 Justifying the whole exercise</b>
12	Principle 8 Pilots are your best friends and your worst enemies
13	Principle 9 Don't forget your checks and balances
13	Principle 10 Big questions are only as good as small realities

## Author

**Mouchka Heller,**  
Autonomous and Automotive Mobility,  
World Economic Forum

© 2020 World Economic Forum. All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, including photocopying and recording, or by any information storage and retrieval system.

1

# Before you even get started



## Principle 1 | Question your assumptions

### Your data is just as subjective as you are

Data is not a magic blanket of objectivity. It carries with it your own biases, prejudices and assumptions. For example, most data collection processes for safety in public transit systems track formal complaints to the police for crimes such as rape and assault. Such processes assume that 1) all victims would formally file a complaint; 2) that all victims would know how to navigate the process for filing a complaint; and 3) those that are not victims of crime feel safe on public transit. If a decision-maker looks at the data they have collected with the assumption that it is inherently objective and presents a complete picture of safety in a given public transit system, that decision-

maker is not getting a holistic picture of safety nor will they be able to present a holistic picture of safety to the general public. Many crimes go unreported, it's a painstaking process to report a crime, and even when you do report one, there's no resolution. There are other examples of crimes that are difficult to include. For example, a bus station defaced with racial slurs may be empty at night but since there are no formal complaints filed, public officials may deem that station safe. Such blind spots can make insights drawn from datasets as valueless as the subjective assessments that were meant to be tested in the first place.

### Hire people who are different from you

Few things in this world are as human as biases and it is humans who are building datasets. Within a team meant to collect and process data to generate insights for decision-making, make sure to hire or include colleagues of socioeconomic

diversity who can ideally challenge each other's prejudices. Data points are only the reflection of on-the-ground realities. Having as many pairs of eyes to test that reflection helps to ensure higher data quality.

### Get ready to be wrong

If you are lucky and do this well, you will generate insights that will challenge your world view, irrelevant of how small or large your project may be. In a project supporting cross-border commuters, a team recently assumed that nurses could be a great focus group because of their unconventional hours and relative inability to work remotely to ease their weekly pace. However, in qualitative focus groups where nurses were allowed

to openly share all their concerns, the aggregated feedback showed that nurses were reluctant to be included in a shared commuting solution because of a preference to keep their overtime hours informal. The entire framework of the operation had to be rethought because of this single point of data, which is the whole reason to engage in a holistic data-driven analysis in the first place.

## Principle 2 | Take inventory, and take it often

### You don't know what data you have until you look

One assumption you have likely already made is how much data you do and do not own. Having a master dataset does not mean you own all the relevant sub-datasets or that you managed subscriptions well enough to be sure of its quality. You also need to pay attention to duplication and

even contradiction. You don't necessarily know how much data your colleagues have collected, how their collection processes have differed, what can be used in combination and what should be rejected out of hand.

## The task of inventory never ends

Just because you have started your process of looking at what data you own and do not own does not mean your colleagues have stopped collecting and erasing data. In some jurisdiction, privacy laws mean that organizations even have set up automatic processes to delete data points

after a certain period of inactivity or on request. Doing inventory manually works, of course, but best practice would be to set up a framework of data integration that is updated continuously and automatically so you know what you have and not what you had.

## Inventory is your most important data

The task of integrating disparate datasets is not trivial, cheap or short. It infers the need for a standardization process that can apply across all datasets, sometimes complex data-sharing agreements, even for internal projects, and an interface capable of not only containing but also

visualizing and ideally synthesizing the data. For projects that necessitate the involvement of non-technical staff, this integrated tool also needs to have user-friendly, front-end capabilities, as well as institutionalized processes for updates and corrections.

# Principle 3 **There is no need to reinvent the wheel**

## Data-exchange frameworks and standards already exist

Numerous cities and organizations around the world have already tackled the question of data-exchange frameworks, setting up global standards that are rapidly gaining in adoption and popularity. With dedicated full-time staff and enormous resources, these organizations still recognize the extraordinary complexity of designing, testing and catalysing the adoption of common data-exchange rules. Public-private data-sharing agreements alone can require full-time work from entire teams for large amounts of time and can end up looking like snowflakes. Cities have adopted largely different stances on data sharing, so the same private

institution also might have different agreements in place around the world and have to manage a separate process for the transfer of data across these regions.

Before diving into your data process and analysis, leverage the knowledge that already exists and rely on data-exchange frameworks and standardization processes that have already been tested and approved in your particular region. A good example of a non-profit leader in this field is the [Open Mobility Foundation](#).

## Data privacy regulations and recommendations already exist

Similarly, do not engage in a guessing game when it comes to data-privacy settings for your product, services or project. Privacy is a culturally sensitive concept that might be completely different from what you would expect from culture to culture and region to region. Rather than go with what you think

is right, establishing focus groups to ask others what they'd prefer, or try and redefine a complex balance, identify the strictest standard in the regions in which you operate and set it as your baseline. It will help with data transfers, adequacy status, public-private partnerships and public engagement.

## Chances are you're not the first to try this

Of course, your idea is unique. Only it isn't. There is a high probability that someone, somewhere, has tried something similar. If that entity partnered with a public or civic organization, there is probably a wealth of information available regarding best practices and lessons learned. Take time to

research similar initiatives to yours and engage project managers. No one is invulnerable to a mistake, misconception, or lapse in judgement. When dealing with an area as sensitive and powerful as data, you don't want to repeat the mistakes of others.

Further, just because you are now engaging with a region that has asked to participate in your project, you should not assume that this region has not already tried similar projects to the one they called

you for. Not only could you learn from the pitfalls and successes of any related past projects but also you could leverage the data they produced to enrich your own and inform your data strategy.

## Principle 4 | **Data likes (selective) company**

### **Reach out to different sectors**

Particularly in an ecosystem as multifaceted as mobility, the complete picture of movement is owned across multiple industries and sectors. Mobility modes are owned by a wide variety of actors, from governmental agencies in multiple jurisdictions to start-ups, and so is their data. In addition, the same user on the same trip, sometimes on the same mode, is tracked through multiple devices that each focus on a different element of that user's movements. An insurance tracker device might register speed and abrupt stops, while a phone would register geolocation, a traffic camera would spot a road infraction, and a CCTV around the corner would collect data on

individuals in the vehicle. If you are trying to get the full picture, for example, of commuting patterns for white-collar workforces driving home from the train station, you might need data from all these sources to get information on a) tendency to carpool, b) safety risks during rush hour, or even c) pressure on drivers in different weather or sharing conditions. A further enquiry to the transit agency could give you numbers on ridership at different stations and transfers to buses versus vehicles parked in the adjacent car park. If your inventory unearths gaps in your datasets, it is worth reaching out widely and forming new partnerships to complete your picture.

### **Combine datasets that can answer the same question**

This does not mean there is value in combining an endless number of datasets but that when you have a single question that is answered either in part or in full by a series of disparate datasets, you need to combine all those datasets following a similar integration strategy to the one for your

inventory to tease out contradictions, overlaps and low-quality data points. In the best-case scenario, you will not only get confirmation of your data but also you will add nuance to your answer. At worst, you will see repetition of certain data points, which can still serve as validation of your work.

### **This might not be the time to be shy about cost**

Data does not have to be expensive, but it can be. If it is meant to be the core of your project, it might not be the item to get on the cheap. There are too many examples of institutions buying cheaper datasets, expecting a certain breadth of sub-datasets and a certain level of quality that they realize, too late, they did not get. If you are able to get the points of information you need without having to pay for them, then great, but make sure you got worthwhile data and not just savings. The same principle applies to integration platforms and processes that ideally need to be automated, easy to navigate, accessible to non-technical staff and verified by legal experts for data-privacy compliance.

and resources. They generally provide small sample sizes that are not sufficient for broader conclusions and applications. Device tracking requires hardware, infrastructure and often complicated partnership agreements. Biometrics tools are legally complex to set up, in addition to necessitating expensive software and hardware. Data is not only a corporate and public asset but also is a currency. Therefore, securing data from partners who have already been through the collection processes is costly as well, however volatile pricing might still be across the industry.

Know what you need, know why you need it, and get exactly that. No more and certainly no less.

Data collection itself is expensive as well. Focus groups take time, space, preparation

2

# While you are at it



## Principle 5 | Data is nothing but insights are everything

### What question are you trying to ask? Why?

You are not trying to create a data set or a pretty visualization tool. The reason why you are engaged in this journey is because you have a question. What is it? Write it on a poster and hang it everywhere around your workspace. Why does this question matter? Why did you need additional data to answer it? What was that data in the first place? What gave you the question to begin with? What will you do with the answer? Who is waiting for it? How will you implement it? How will you support it?

Data is nothing, just a group of points of information that may or may not be correct or valuable. The insights you derive from that data are the foundation of the decision you are about to make; of the action you are about to take. The first step to move from a data analyst to an insight generator is to always keep in mind what it is you are asking and why.

### Cut through the noise

First and foremost, having a clear list of questions you want to answer allows you to cut through the noise. It's difficult to find a dataset that will perfectly answer your question. In fact, you will likely receive a dataset with multiple interesting points of information that will get you lost by suggesting

different answers from the one to the question you are asking. By remaining focused on the reason why you are on this journey, you can cut through the noise and identify the few data points that actually matter. It will save you time, money and energy, and keep you on task.

### Focus on the user

Second, whatever project is justifying your foray into data is only relevant because of the impact it will have on the user. Any piece of information that does not serve the user's bottom line can be

discounted as noise. Any question you are focusing on, any answer you are trying to find and any narrative you are trying to build should focus on the user as well.

## Principle 6 | Less is most definitely not less

### The cost of too much data

Restrain your team from creating too many datasets and dealing with too many different sources of data beyond what you have collectively agreed upon as necessary. Too much data creates challenges with standardization and integration, with quality and management, and with budgeting. Beyond prioritizing smaller data pools, you should implement a single standardization and data-

exchange framework for everyone to follow, from high-level policies to the weeds of data operations. As discussed in Principle 4, data is expensive. The time of your team is even more precious. Too much data ends up looking like a Tower of Babel, a wonderful visual of diversity and curiosity and a place where no one can talk to anyone else.

### Learn the difference between critical vs hollow gaps

Not all gaps are created equal. To go back to the commuter analysis example, it might be relevant to push your analysis as far as make and brand of the car, perhaps to help establish the socioeconomic class of different commuters to qualify their behaviour, but colour of the car or haircut of

the driver is likely something you can afford to ignore. Going one step further, if you already have information on the average socioeconomic class of the residents at a certain stop versus another, you might be able to rely on simple usership data between these two stops and not bother with the

data related to make and brand of the car, which can itself have more variables than income or class. Something you thought was a critical gap can

actually be a hollow gap not deserving of your time and resources.

## Don't be scared of qualitative judgement

Data is not the purely quantitative game some make it to be. Your original questions, assumptions, hiring strategy and partnership management, which are all qualitative, inform insight generation. Your qualitative judgement can

also allow you to assess the level of quality of the data you are working with, identify pitfalls and manipulate your data differently. Qualitative and quantitative thought processes are best as allies in your data journey.

## Principle 7 | Ambition has many traps

### Don't try to do everything

Data is powerful and awe-inspiring. It is easy to see endless potential in the datasets you start having access to, once you get started. However, trying to solve everything, like trying to transform an entire mobility system, is the best way to accomplish nothing. Working with data is actually a granular, highly operational and technical, bottom-focused job. It is quite different from asking lofty questions in a philosophy class and relying on historical patterns or cultural references for argument. The size of datasets is often in terabytes, containing millions of points of information that necessitate dedicated software, hardware and teams to properly sift through. Patterns in this abundance of data do not come packaged but require deep thought about parameters and purpose, and often necessitate their own decision-making process.

For example, in a recent data research project by Southeast of Michigan Coalition of Governors on accessibility to public transit, the research team had to decide what accessibility meant and how to define impact. At a high level, accessibility can mean anything from wheelchair ramps to the affordability of transport, but that is not a definition that can be used as a basis for data-based research. First, the team had to limit their definition to the capacity to reach a public transit stop. The

next obstacle was the spectrum of vulnerabilities in a given population. Would a public transit stop without a bus shelter be considered accessible in winter? Should a stop on top of a hill without a pavement be included? How do you account for such parameters when manipulating your dataset? After weeks of data manipulation, the team settled on a definition of accessibility based on time. If a given person could technically reach a bus stop by walking for up to 30 minutes, that bus stop would be considered accessible.

The same question applied to impact. Most thought leadership on accessibility emphasizes disconnection from core services as the main impact. How can that be quantified and tracked? The team eventually settled on seven core destinations that did not include things like therapist offices, libraries or police stations, all services that can become essential at different stages of life and for different households.

Does this make the research valueless? Of course not. It means the data could be processed with some level of rigour and that the study can now be leveraged as a starting point for a) qualitative judgement and b) further questions and probing.

## Ask one question at a time

In fact, even for ambitious projects, best practice is patience, asking one question at a time. Can we test for accessibility at all? Is time a good measure that can work across our datasets? Can we show blind spots and discrepancy in connection to core services? All these questions were answered by the Michigan study, one by one, so the research could proceed. The stage would be to ask questions like, at what point in time do inequities in accessibility disappear. (Is everyone within 30 minutes of a public transit stop? One hour? Two days?) Are

there core services that are more evenly distributed than others? Can that be attributed to specific neighbourhoods that can be correlated with targeted socioeconomic classes, race, gender or ethnicity?

There is no limit to the depth and breadth of the ultimate insight to which your research can lead you, but the road is paved with single questions that need to be addressed, explored, quantified one at a time for you to arrive to a defensible result.

## Don't ignore unconventional resources

Critics of the Fourth Industrial Revolution argue that its access is limited to the rich and powerful, which applies to public and private institutions as well. The cost of data discussed earlier is certainly prohibitive to many, in terms of financial means, time and human capital. However, that

does not mean these best practices can be observed only by the elite. Far from it. Instead, smaller-sized cities, countries and organizations should look to less conventional, often forgotten resources like students to fill the gaps.

3

## Justifying the whole exercise



**Pilots can be a Bermuda Triangle**

If you have done everything right up until this point, you now find yourself with a problem statement to a real-world problem, solid insights based on high-quality data, and you need to test these insights back in the real world. The greatest trap to your project at this point is to create a pilot just for the sake of demonstrating your data, instead of making sure you have all the conditions necessary for a

true test. In fact, this pilot is meant to act as just another data point in your research. If you forget that, it will gain a life of its own and can become a Bermuda Triangle that sinks all your efforts at the bottom of the ocean, where no one will ever find them again. Instead, all focus will be on that pilot, which itself won't be properly informed by your insights anymore.

**Pilots are tests, not dictators**

Instead, pilots should function as tests of your insights. They need to be carefully designed so they occur in conditions that properly reflect the parameters set in your data analysis and can feed information back to your insight generation

process. A pilot should not be launched for the sake of it. It should not be launched in Region X because that is where the funding is or where your boss has a great connection. It exists to serve the data journey, not the other way around.

**Bad pilots give bad data and biased confirmations of your hypothesis**

The greatest danger, of course, is that bad pilots will feed you bad data, which will lead you to the wrong insights and decisions. To go back to the example of commuters driving home from train stations – some suburbs that see such commuting patterns host lower-income households where the culture might be to own cars for each head of household even if that car is a major financial burden on the family. Let's say a decision-maker wondered whether implementing free transfers from the local train to the bus system would encourage residents to use shared, higher occupancy modes of mobility. That decision-maker receives federal funding to subsidize transport in lower-income communities within their specific state and environmental organizations further encourage them to spend these funds in lower-income communities that still have high rates of car ownership.

If the decision-maker went through our earlier questions of considering whether people already used carpools, how many riders parked and used their car or took the bus from the train, what safety hazard that might create, and how much additional insurance potential speeding cost the commuters

only to launch a pilot in that neighbourhood of low-income households with a proud history of car ownership, they would lose all the insights they gathered with their data. They would launch a pilot for free transfers and that would not be enough to convert residents from a long-standing culture and point of pride. The worst that could happen would be to deduce that financial incentives do not impact ridership of shared mobility solutions. It did not, in this particular case because the pilot was determined by variables external to the tests needed for the data, not because the deduction is correct. On the other hand, if the decision-maker had found a part of the region where insurance premiums went up within a year of car ownership, largely due to speeding during rush hour from commuters driving home from the train, added to high parking rates and regular fines for road infractions, within a lower-income community that did not present the cultural factors of the first one, then they could run a campaign focused on cost-saving and observe whether insurance premiums and fines were good enforcers of the incentive they had created. In this example, the pilot only exists to serve the data, not the other way around.

## Don't forget your checks and balances

### Don't assume privacy is just about the law

Privacy is a highly culturally sensitive and personal issue. The case study of Alphabet's engagement in Toronto is a great example of how data privacy regulations are not always good enough parameters to determine perception and reaction

from the public. Data privacy sensitivities are excellent checks and balances to calibrate your risk throughout the data collection and analysis processes.

### Invite the party pooper in

The old principle of surrounding yourself with "no men" applies even more to data-based initiatives. You need someone on your team who will look at whatever you put in front of them with scepticism,

spot gaps and potentials for errors and point it out. Respect this person's view and don't get defensive. If they can poke holes at your data and insights, it is likely many others can.

### Set up an overarching framework as your frame or reference

Another key best practice to ensure checks and balances is an overarching framework, both technical and conceptual, to help you remain on track. Once you have selected your data-exchange rules, do not change them unless it is essential to the integrity of the project. Once you have committed to a certain level of data privacy rules, do not walk it back once it gets hard.

making, then keep it oriented towards concrete action that is in tune with the realities, funding, needs and governance gaps of the region you are interested in. If your goal is to create a new avenue towards profitability, have enough courage and honesty to stay the course if the data shows you that the project in its current state is not financially profitable. That honesty will allow you to identify true drivers of profitability and debunk mistakes that could have led to a premature end of the project.

Perhaps even more importantly, know what you are working for. If your project is about public policy-

## Big questions are only as good as small realities

### You won't know what you don't know

At the end of your project, you will remember most vividly every time you were proven wrong and realized how much you didn't know. When you know nothing about a topic, it is easy to believe you know everything. Once you start digging, you know you are on to something when you realize how much it is you don't know.

With data work, it can be hard or impossible to know in advance where your gaps will be. You may not realize you will need a specific sub-dataset, or that you are not aware of a statistical software package that could have made all your efforts go exponentially faster. You also don't know how good your data is until you get into the thick of it, or if your hypothesis has any chance of holding until you are halfway through demonstrating it.

Therefore, embarking on data work demands a high level of comfort with ambiguity. You still need to be able to formulate a frame of reference, a problem statement, to choose a set of data-exchange rules and standardization processes. You will have to explain your projects a million times to a million people in a million different ways to get the support you need to get it started. You probably will need to apply for funding by predicting output. Yet, you won't know just how on or off target you are until you are deeply committed. However worthwhile your original question, vision and hypothesis may be, the reality of the data you have available to you might kill them.

Be ready to adapt, get comfortable with ambiguity and accept that you might have to walk away.

## What seems insignificant could be your game-changer

In Principle 1, we presented the example of a commuting pilot for nurses, which hit a roadblock when the team realized they did not know the role that informal overtime hours played in the position and professional life of a nurse. The insight itself came from a question in a focus group that asked respondents what form of communication they would prefer to notify them that their shared shuttle had arrived. The project managers were first alerted when only passive communication channels were checked. This apparently insignificant piece of data led the project managers to further probe respondents, resulting in the eventual discovery that some respondents preferred to be able to miss boarding the shuttle altogether. Further probing, led informally, led to the realization of the overtime

parameter, which had not been considered earlier. That piece of data changed the entire pilot.

Problem statements, frames of reference, ideas for policies all remain conceptual and intellectual exercises. Yet, they touch the reality of riders in a very concrete and granular manner. It takes art and tremendous effort to climb down from the conceptual to the real and be able to go back up. However, good data work is done by project managers able to embark on that specific trip multiple times in the duration of a project. That is how project managers are able to catch apparently insignificant pieces of information, lead the deeper probing and come up with the adequate level of insight that transforms the project appropriately.

## Value is measured by impact, not prowess

In spite of its newly gained reputation, working with data is not a sexy, fun or easy endeavour. It requires a lot of humility, patience, comfort with ambiguity, attention to detail and collaboration along with ambition, audacity, instinct and resilience. Data project managers are not the Olympians of our time. The best of them are not on top of an inaccessible mountain, reigning over the rest of the world as gods. Instead of prowess, the best of them drive impact.

Posing as a problem statement the complete transformation of a regional mobility system, for example, is seeking a demonstration of prowess that will take time and effort. Consider this problem – new mothers are unable to attend their newborn check-up appointment which leads to infant deaths. These mothers live in a region where climate, disconnectivity and a cash-based economy prohibit the use of most new

mobility technologies. The resulting pilot might be as humble as a bunch of closed bus shelters strategically placed near hospitals and at key intersections of lower-income neighborhoods, but the result will include saved lives. It's not a glamorous solution, but it's impactful.

There is not as much glory as you might think in data work. In fact, if you are a good data project manager, you will spend most of the time in the shadows, with your neck and back hurting from spending too much time in front of a computer. Don't enter because you want to be celebrated. Leave the work of operationalization to the full-time professionals and concentrate on asking the right questions to the right people, on treating the resulting insights with integrity and on driving real change. However humble it might look on the outside, it will produce impact and eventually that is what will make you shine.



---

COMMITTED TO  
IMPROVING THE STATE  
OF THE WORLD

---

The World Economic Forum, committed to improving the state of the world, is the International Organization for Public-Private Cooperation.

The Forum engages the foremost political, business and other leaders of society to shape global, regional and industry agendas.

---

**World Economic Forum**  
91–93 route de la Capite  
CH-1223 Cologny/Geneva  
Switzerland

Tel.: +41 (0) 22 869 1212  
Fax: +41 (0) 22 786 2744  
contact@weforum.org  
www.weforum.org