Human Sensemaking in the Smart City: A Research Approach Merging Big and Thick Data

ANNELIEN SMETS
imec-SMIT, Vrije Universiteit Brussel

BRAM LIEVENS
imec-SMIT, Vrije Universiteit Brussel

This paper aims to contribute to the debate on the integration of ethnography and data science by providing a concrete research tool to deploy this integration. We start from our own experiences with user research in a data-rich environment, the smart city, and work towards a research tool that leverages ethnographic praxis with data science opportunities. We discuss the different key components of the system, how they work together and how they allow for human sensemaking.

THE BIG-THICK CHALLENGE

Both in industry and academia, we witness an increasing focus on human behavior and individual experiences. In this era of cheap data storage, this interest translates into the collection of as much as data as possible of someone's actual behavior. In digital environments for example, each keystroke and each click are precisely being logged, creating high volumes of data. By means of different computational methods these digital traces are then being transformed to gain a deeper understanding of users' behaviors in order to improve products or services. With the emergence of the Internet-of-Things (IoT), the increasing number of physical objects that are being connected, we observe the same phenomenon in environments of our daily lives that originally were not perceived as digital.

Within urban environments, especially, this connectedness is already resulting in an abundance of available data originating from sensors deployed throughout the entire (smart) city. However, the complexity of the urban environment adds additional challenges to the process of transforming data into meaningful information and actionable knowledge.

First, one needs to know what to measure. This includes safeguarding the balance between predefined research variables and the level of serendipity, as well as identifying what should be measured at which moment in time. Second, the entanglement of multiple domains in the city, results in a high variety of data streams that require different expertise knowledge. We need a thorough and unambiguous understanding of what is measured and what this exactly means (Hey et al. 2009). For example, air quality is often represented as the amount of parts per million, but for a layman this is meaningless data.

This links up to the third challenge: contextualization. A great deal of the existing data originates from individualistic systems that have no sense of the collective: "it might be big, but not very useful unless it is set in a wider context involving other data" (Batty 2016, p.323). This means that, in order to really understand the data, we do not only have to look at the data itself, but also take into account its broader context. To stick to the air quality example; to interpret the data of an air quality sensor, one has to take into account

parameters such as traffic density and weather conditions, as they influence the measurements and therefore are important to get an understanding of the situation.

The fourth and final challenge, we call it human sensemaking, relates to providing an additional layer to the data, which captures what this data means to the citizens. Again, in case of air quality, we want to add a human interpretation to each of the different quality levels and identify if these interpretative categories align with the different categories resulting from the knowledge domain; is "low air quality" actually perceived as unpleasant?

This human interpretation is necessary as it contributes to a better understanding of experiences and human behavior. Similar to the example of using digital traces to improve services, we need these insights to address urban challenges. Smart cities aim at improving the overall quality of life in the city and following the progress made by merely technological solutions, we are now heading towards smart city solutions that target a change in human behavior (de Oliveira et al. 2015). However, to design smart city solutions that are able to bring about this change, we need to connect the data resulting from the city with in-depth insights of people's behaviors and experiences. This means that we are now dealing with situations where we have to determine if citizens are aware of the smart lightning system in their street, study how they experience a smart traffic light or examine how safe they feel on the road. Can we infer this information from the data the smart city is providing us? And more importantly, how can we do this?

In this paper, we address this sensemaking challenge by exploring the creation of a research tool that leverages ethnographic praxis with data science opportunities. In this way, we aim to contribute to the current call from research to provide actual tools to analyze and work with data (Churchill 2017). The desire to integrate these two disciplines is not only motivated by common interests of ethnographers and data scientists (Curran 2013), but also by the additional value generated by integrating both disciplines: large-scale in-depth insights. While an ethnographic approach allows for an in-depth understanding of people's beliefs and behaviors in their context (thick data), data science is able to detect patterns in data points collected at a large scale (big data) (Wang 2016). Although most of the discussions on this integration remained conceptual, they have already led to numerous valuable insights. Consequently, this paper mainly focuses on the research tool that we are currently developing and the main insights from the research leading up to this.

TOWARDS HUMAN SENSEMAKING

The term "sensemaking" knows many interpretations and, ironically, in its broadest sense it could be described as "getting an understanding or attributing meanings to something" (Kari 1998). In our sensemaking approach, we start from Dervin's assumption about individuals' sensemaking in the way that "they experience and observe their world differently and need to create meaning or make sense of their world" (Dervin 1992, p.62). The goal of our human sensemaking approach is to being able to capture these different senses (let it be in the forms of meanings, experiences, etc.). More specifically, within our smart city research, we use sensemaking to refer to the process by which a participant (most often a citizen) gives meaning to his or her experience related to the interaction with a service, technology or an urban environment.

Dervin argues that sensemaking can be considered as "behavior, both internal (cognitive) and external (procedural), which allows the individual to construct and design his

or her movement through the time-space context" (Du 2014, p. 29). In order to understand this behavior and its outcomes we need to be able to study it. This comes, however, with some challenges. First, we need to know the context in which this behavior occurs and second, as it is time-space dependent, we need to be able to capture the experiences throughout different contexts.

The research tool presented in this paper, addresses these challenges by bridging the gap between big and thick data and thereby enables studying experiences within their context. Not only does the tool reinforce the relationship between big and thick, it also adds meaning to each data type individually by providing a greater context. To this end, the research tool relies on four sensemaking strategies, which arose from previous experiences in working with big and thick data and the different needs we identified: time-space dependency, pattern detection in large data-sets, gathering subjective experiences and combining those with objective data to create meaning. We define the following sensemaking strategies:

- **Contextualization**: as Rato (2000) pointed out, being able to determine the 'right' context within which to situate one's analysis, is one of the key elements in both ethnographic and sensemaking research. This strategy is in line with the Living Lab contextualization phase as defined by Pierson & Lievens (2005) and the grounded theory approach. In our approach, contextualization relies on various big data collection methods (such as logging data, tracking data) in order to get a profound insight in one's behavioral context.
- **Semantics**: the semantic strategy builds further on these collected datasets and provides a first meaningful layer to the gathered data points. By means of a semantic framework, raw data is being translated into meaningful data objects. For example geolocation data is being translated into amount of times being present in a certain place, the duration of a person at a given spot.
- **Analysis**: similar to the semantics, the analysis strategy supports researchers to get meaningful information out of the data points. This strategy allows to detect patterns in large data-sets as well as summary statistics that allow researchers to put the data into perspective. Another activity in this strategy is the identification of user profiles (i.e. clustering groups of users).
- **Human interpretation:** whereas the previous strategies are still a construct of the research, the human interpretation strategy allows for a direct sensemaking process by the subject itself within a particular context. By means of using experience sampling techniques it is possible to capture the experience at the time and space of its occurrence. This strategy thus allows to add another additional layer to the data consisting of experiences, emotions, motivations, etc.

These strategies could broadly be divided in two categories where one of them is more concerned with dealing with big data whereas the other deals with gathering the thick data. Nevertheless, these categories are non-exclusive and there is a continuous interaction among them. A specific example of the realization of these strategies and how they work together is provided later in this paper.

In the remainder of this paper, we describe a research toolkit that allows to put these four sensemaking strategies in practice. We position this toolkit within a Living Lab research approach, since this is our primary way of working and hence the requirements of the toolkit find their origin in our previous experiences with Living Lab research.

THE URBAN LIVING LAB

The increasing focus to study human behavior in urban environments, strengthens the need to perform contextual research at large scale and integrate big and thick data within smart city research. At the same time, we observe smart cities having the potential to actually accomplish this integration: their vast amount of available contextual data allowing for exploratory data analysis, reinforces the ethnographic approach by evoking a greater level of serendipity (Rivoal & Salazar 2013). Access to this data and means to turn this into information are obtained by perceiving the city as a Living Lab (Coenen et al. 2014). The city, as an urban innovation ecosystem, thereby acts like a dynamic open experimentation environment involving its citizens (Veeckman & Van der Graaf 2015). Juujärvi & Pesso (2013, p.22) define the Urban Living Lab as follows:

> "a physical region in which different stakeholders form public-private-people partnerships of public agencies, firms, universities, and users collaborate to create, prototype, validate, and test new technologies, services, products, and systems in real-life contexts".

By approaching the smart city as a permanent Urban Living Lab, it internalizes ethnographic characteristics such as naturalism, understanding and discovery and enables studying users' behavior and experiences in the wild (Pierson & Lievens 2005). Additionally, the Urban Living Lab approach provides a framework to co-operate with community partners and establish a method to cross-validate the observed patterns in the data (Kontokosta 2016) as well as to complement these with their human interpretative counterpart.

These characteristics are of big importance in studying the sensemaking process and its outcomes. First of all, the ability to study behaviors and experiences in the wild results in a more genuine sensemaking process that is not (or at least less) influenced by the research itself. Additionally, the Urban Living Lab provides a semi-controlled environment, which allows the researchers to take into account the time-space context of the observed behavior and experiences. At the same time, this time-space context can also be used as an entry-point to steer parts of the research (e.g. only ask questions when the subject is in a particular context). Moreover, the Urban Living Lab allows for big data collection exceeding the bounds of the recruited participants, while it also facilitates the collection of thick data by having dedicated interactions with the set of participants.

Urban Living Labs do not only hold great promises to facilitate the integration between big and thick data, they also need tools that allow for this integration. Currently, we notice two main problems that would benefit from it. First, there is what we call the problem of unavailability. Within the current Urban Living Lab projects, we notice that there is a lack of objective data. To gather data about the interaction with a digital service, researchers can (and do) rely on logging data. However, with regard to the context or the actual behavior of citizens in the urban environment, researchers need to rely on reported data by the participants. This often done by means of traditional qualitative methods such as observations, interviews, focus groups and diaries, whilst -especially in the smart city context- there are possibilities to directly obtain this contextual and behavioral data.

The second problem is more related to the actual integration of big and thick data and can be described as the problem of *asynchronicity*. By asynchronicity, we mean that there is

data available on the behavior or the context of the citizens, but this data is only thickened some days (or even weeks) after the behavior occurred. This results in a significant delay between the generation of the data and the moment at which it is being interpreted and gets a meaning. Although this time span can have beneficial outcomes supporting self-reflection, it does also result in a potential recall bias and the experience might be distorted over time. The latter could be due to repeated experiences, but also due to external influences. In our opinion, the time delay results in post-experiences rather than experiences in the moment. These post-experiences are definitely relevant as well, however, within our research scope, we are looking for thickening strategies within the moment itself.

Having observed these two main problems in the last years, we recently piloted some case studies where we tried to augment current Urban Living Lab practices by using digital trace data and at the same time overcoming this problem of asynchronicity. In the next section, we will briefly describe two cases and pinpoint our lessons learned, which have been the major drivers to continue our work on human sensemaking in the smart city. Our experiences with these case studies have also greatly defined the requirements for the research tool that is described in this paper.

Thick Understanding in the Smart City

The two case studies that are described in this section are part of City of Things[1]. This is a Smart City (Urban) Living Lab and IoT testbed located in Antwerp (Belgium) that aims to bridge the gap between the quadruple helix (government, research, industry and citizens) by bringing them together and let them collectively test and validate innovative solutions within a real-life environment (Latré et al. 2016). Currently, different projects are ongoing within this Urban Living Lab focusing on mobility, air quality and traffic safety. Except from the various technological challenges in these projects, as user researchers we are challenged to apply proper methods and tools to capture the needs, requirements and the real-life experiences of the various stakeholders (mainly citizens) to gain sufficient insights to steer and evaluate the design process. The two projects that we will describe, have been set up to explore the possibilities of the technological infrastructure in City of Things and to what extent this can benefit our user-centered, ethnographic research steps. The first project, Citizen Bike, mainly focused on how we could use big data in our research process, while the second project, Be-Und, explored the use of contextual data to trigger in-situ interactions with the participants.

Citizen Bike - In this use case we studied how the cycle experience in the city could be improved. In our approach, we first wanted to identify the current cycling experiences throughout the city and our goal was to do this by means of geolocation data (to capture the movements) and data from different sensors in the city (to capture the context) (see Boonen & Lievens (2018) for a thorough description). Instead of using an off-the shelf smartphone application that tracks one's location and movements, we chose to develop a customized device that allowed for continuous tracking and provided real-time data without any intervention of the end-user (Figure 1). This device also contained two buttons that allowed participants to share their experiences while cycling. The use of this device enabled to gain a deeper insight in the experiences of the cyclists by combining big and thick data methods.
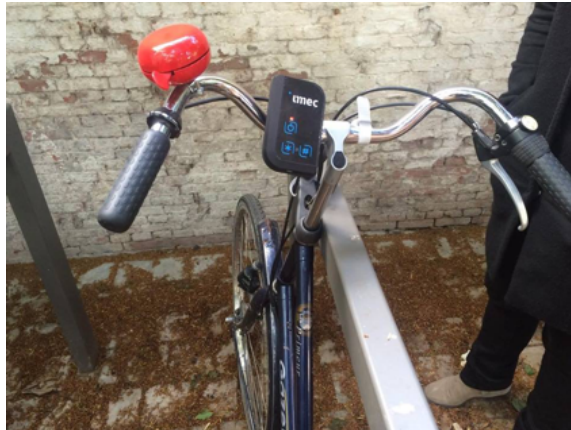
Figure 1. Citizen Bike device mounted on a user's bike.
© Boonen & Lievens (2018), used with permission.

During the field study the device provided us two types of data in real-time while the participants were cycling throughout the city: a continuous stream of sensor data and user experience data (as the users were asked to push one of the buttons when they experienced a positive or negative situation while cycling). We analyzed these data points and integrated some of them (Figure 2). Although this analysis provided us some general insights, it lacked subjective data from the respondents. Subsequently, based on these first insights, we performed additional qualitative research to get a deeper (and hence ticker) understanding of their cycling experiences.

Figure 2. Visualization of users' mobility (left) and their cycling experiences combined with different levels of noise (right).
© Boonen & Lievens (2018), used with permission.

The Citizen Bike case was hence one of our first attempts to use digital data from sensors to capture one's experience and make meaning of it. In terms of methodological approaches, one could say that we used the experience sampling method to capture the cyclists' experiences, because we allowed them to indicate their positive or negative experience in the moment by using the interaction buttons. As mentioned before, we used these data points as entry points for our qualitative interactions with the participants and this turned out to be a very valuable approach. As mentioned in Boonen & Lievens (2018, p.210):

> "(...) this helped us to stimulate users' recollection. For instance, as soon as we confronted one participant with a certain push on the button, she remembered the whole route and could provide more information on why she pushed the button whilst reflecting on her general use."

Being able to present the captured data to the user turned out to not only be useful to get a deeper understanding of the experiences they did share, but it was also helpful to identify why they did not share something. For example, the noise levels at the location of the participant might have been very high, yet this situation was not marked as a negative experience; why not? This insight taught us that having this continuous stream of data on contextual and/or behavioral events is important to allow the researcher to look for interesting situations or patterns by him- or herself, rather than solely relying on the input of the participant. Moreover, during the qualitative research steps, we also learned that being able to interact with the participants in real-time (at a greater extent than interaction buttons) would have been more beneficial because the real experience and the subjective feelings that are associated with it, are often hard to recall and this could have been overcome by being able to interact in the moment.

Be-Und - The objective of this second use case (see Smets et al. (2018) was to explore how we could map and understand current behaviors of citizens, as a first research step within a broader behavior change research approach as defined by Spagnoli et al. (2017). To study the use of digital methods and objective data on both the behavior as well as its context, we first developed a generic behavioral understanding method, which we then applied to a specific use case: commuting behavior.

Whereas the Citizen Bike project was able to capture cycling behavior (route), some contextual parameters and a thin, in-the-moment description of the cycling experience (positive or negative), this use case aimed for a greater in-situ interaction with the users to explore their experiences. To this end, we set-up a research design that relied on the context-aware experience sampling method, which improves the traditional experience sampling method "by using sensing technologies to automatically detect events that can trigger sampling and thereby data collection" (Massachusetts Institute of Technology 2008). This implied that we, as researchers, were able to interact with the participants and ask questions about their experiences based on their context.

This approach us to properly investigate which contextual factors influence one's decision to take either the car or the bike to go to work. For a two-week period, we equipped 6 citizens of Antwerp with a context-aware experience sampling tool that allowed us to track their location. This data provided us information when they are leaving at home and when they arrived at work. This tool sent a notification to the researcher's smartphone when the participant arrived at a particular location (in this case work). This allowed us to immediately, and hence still in the moment, send a tailored questionnaire to the participant to acquire a thick description of the commute and what influenced the commuting decision (car or bike).

To enrich and validate the collected data, an additional focus group was organized after the two-week period, to confront the users with the data and insights. Moreover, in contrast to Citizen Bike, the ability to interact with the participant in the moment (i.e. when they arrived at work) turned out to be particularly interesting since we were able to inquire really specific details that might be forgotten when time has passed. Nevertheless, we also

experienced some difficulties with the actual implementation of the context-aware experience sampling and listed some requirements for a future tool: real-time accessible data, integration with contextual data, user-friendly content management system, pre-scripted triggering rules and question sets that can be send automatically and to each participant individually (Smets et al. 2018).

CITIZEN TOOLBOX

Taken into account the learnings from the above cases (Boonen & Lievens (2018) and Smets et al. (2018), we designed a toolbox that would help us to overcome some barriers and hence satisfy our need to be able to conduct user research within an Urban Living Lab while taking advantages of the available digital contextual and personal data. The main goal of this research tool is to provide user researchers an easy and accessible instrument to collect objective, big data and use this data to interact with the end-users to make sense of this data by combining it with thick data. We first present a high-level description of this tool indicating the key concepts and data flows. Next, we dig deeper in the specific components, their configurations and usage. Thereafter, we describe four ways in which the Citizen Toolbox encourages sensemaking.

The development of the Citizen Toolbox was approached by means of an iterative and agile design process (including user story mapping, prototyping, etc.) involving all stakeholders within our organization (user researchers, project managers, database managers, software developers and user involvement coordinators amongst others).

The tool allows to unlock various contextual and personal data streams and, based on these data, provides the ability to directly interact with participants to gain a thicker understanding about their behavior and experiences. More in particular, it consists of (1) a personal tracking system that monitors someone's behavior, (2) a database containing contextual sources (such as weather and traffic) and (3) a rule-based experience sampling method that enables contextual inquiry based on the input of (1) and (2). The main strength of this set-up is that it allows to perform contextual research at a large scale and to integrate big and thick data, thereby augmenting the overall value of the data. Next to this overall research goal, we also identified some additional requirements that we perceive as crucial for the Citizen Toolbox to be used as a tool to support user researchers (Table 1).

Table 1. Requirements and corresponding features in Citizen Toolbox.

Requirement	Feature
Short time-to-research	Fixed set-up and limited scoping Pre-configured devices
Independent of technical development	End-user programming Library (data sources) Standard reporting
Reusable in different projects across different domains	Library (rules, data sources) Standard reporting Stream framework Multiple data input modalities (plug & play)
Safeguard serendipity	Flexible system configuration
Activity monitoring	Various logging data

For the design of the Citizen Toolbox we chose for a solution that consists of a pre-configured set-up in terms of devices and database configurations. This does not only allow us to roll it out quickly, but also allows the tool to be managed and configured by user researchers rather than developers. Moreover, there is the need to safeguard serendipity as it is inherent to ethnographic praxis. Consequently, the Citizen Toolbox allows for flexible system configurations, meaning that not everything should be defined upon the start of a project, rather a researcher can still make changes during the project based on empirical evidence.

> "Ethnography is an inductive science, that is: it works from empirical evidence towards theory, not the other way around. This has been mentioned several times already: you follow the data, and the data suggest particular theoretical issues." (Blommaert & Jie 2010, p.14)

Building Blocks

A schematic representation of the research tool can be found in Figure 3. In general, one could identify three major types of components, which we will extensively describe below: big data input, processing components and thick data capturing.
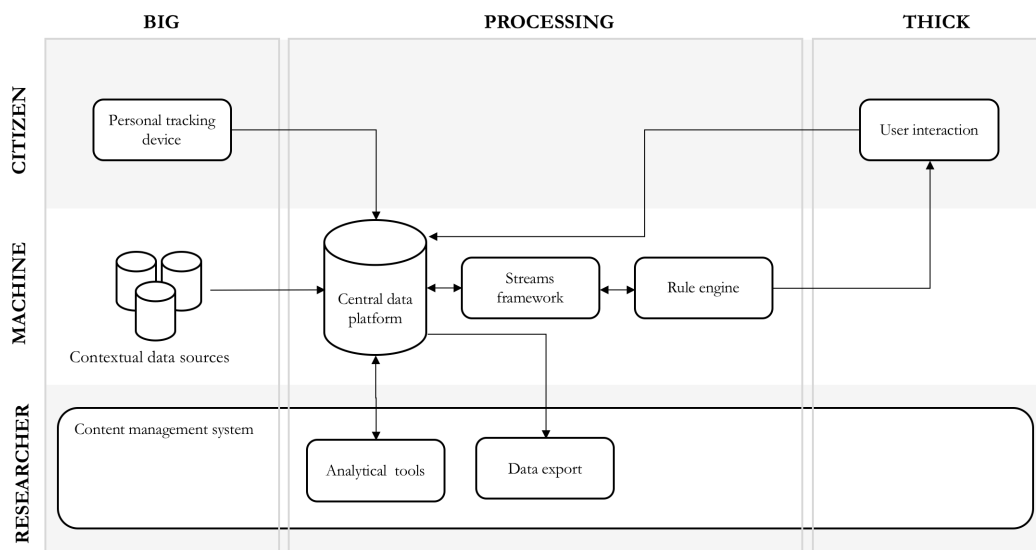

Figure 3. Schematic representation of the different components of the Citizen Toolbox.
© Authors' illustration.

Big data input - The Citizen Toolbox foresees two big data sources: a personal tracking device and contextual data sources. The former allows to capture behavioral data, which could be one's location, movements or physiological data such as heartbeat rate or galvanic skin response. In the first version of the Citizen Toolbox, the personal tracking device keeps track of the geolocation of the participants, using a GPS sensor. This is our crucial

component to be able to identify the context of the participants because their geolocation can be connected with contextual data from external data sources.

These external, or contextual, data sources allow to gather additional data on contextual parameters. To this end, we foresee a library of various existing, external data sources that are integrated in the tool and can be used in a specific research project. For our research within the smart city, we think of existing external data sources such as weather data, air quality, traffic, number of available parking lots, etc.

Processing components - The big data input is centralized to a central data platform. This platform automatically takes care of the required data processing activities (e.g. extracting the data from its source, cleaning, transforming into the correct data types and loading it into the database). In addition to the big data input sources, the data platform also stores all inputs from other components in the system.

Next, there is the streams framework responsible for the semantic translation of the raw data into data objects that are meaningful to the researchers. The main rationale of this component is the idea that we need some sort of aggregated variables to continue our sensemaking process rather than the individual data points coming from the tracking device and/or external data sources. For example, we want to learn from the GPS data when someone is at home, however, this geographical point 'home' will contain multiple coordinates which we are not interested in, but only in their aggregated variable 'home'.

The rule engine is the core processing component. It is responsible for the automatic handling of predefined actions. These actions could be diverse like writing variables to a database or initiating the calculation of an aggregated variable. This kind of actions are however rather useful in practical terms and what we consider to be the most valuable functionality of this rule engine is its ability to trigger questions to the user. The triggering of these questions is hence based on meeting a condition consisting of one or more variables. At the moment this condition is satisfied, the specific question is sent to the interaction module (see below) and the response given by the participant is stored in the data platform.

Finally, the Citizen Toolbox also contains a content management system (CMS) that allows researchers (and other stakeholders such as project managers and panel managers) to configure the system, view the data and monitor the project's current state. The two most important modules of the CMS are the analytical tools and data export tool.

The analytical component represents a set of analytical tools that allow the researcher to analyze the data. The tools allow the researcher to generate some first, standard insight reports, that act like a dashboard for the researcher. Additionally, these tools provide the ability to perform profound statistical analysis of which the results can be re-inserted in the central data platform. Apart from the descriptive analysis, the main focus will be on the detection of patterns and deviations, as we want to be able to thicken these insights with end-user input, through the experience sampling or as input for more in-depth qualitative research.

At last, the data export module allows researchers to download raw data files, but also provides the possibility to download standard reports.

Thick data capturing - Together with the rule engine, the interaction module represents the core of the thickening capabilities of the Citizen Toolbox. After all, this interaction module allows to add an additional layer to the data with the users' interpretation and/or

experiences. This interaction with the user is supported by a smartphone application that is installed on the participant's own device. The choice to install this application on their own device rather than on an additional device that is provided by us, is deliberately made based on our previous experiences with user research. After all, our goal is to minimize the response time (i.e. time between sending a question and receiving a response) in order to maximize the number of in-situ interactions. Therefore, we chose to install the interaction module on their own device as they are more likely to see a notification on this device compared to a device that they only have to use for the purpose of the research.

This smartphone application sends a notification to the participant when he or she receives a question (triggered by the rule engine). The participant is then able to answer it in the application.

Four Times Sensing

So far, we have described the different components of the Citizen Toolbox and their main functionalities. However, we still need to demonstrate how the configuration of the toolbox allows for the four sensemaking strategies as described above. Figure 4 represents these sensemaking strategies and how they are enabled in the Citizen Toolbox configuration. To illustrate them, we will briefly describe a case study and highlight the activities that allow to thicken the data.

The Story - A high-school located in the center of a large city wants to stimulate its students to come to school by bike. A first survey among their students indicated that first-graders are less likely to bike to school than older students. One of the major barriers that these first-graders face, is the traffic near the school, which makes them feel unsafe. The school wants to support these first-graders to take the bike to school, but they also want to get a better understanding of other cycling motivations of other students because they feel that getting such an understanding might benefit all students in the long-run. Given these needs, a research project was set up, involving the use of the Citizen Toolbox.

Contextualization - The first strategy allows to integrate multiple data sources and consequently makes it possible to see the data in its wider context. In the school example, this could be the integration of the GPS data from the personal tracking devices and information on the traffic (that could be an indicator of road safety). However, the mere integration of these datasets could not be very useful yet, although, in the Citizen Toolbox, it can be directly used as input for other sensemaking strategies to turn it into actionable insights.

Semantics - The semantic strategy heavily relies on the streams framework since this component is responsible for the semantic translation of the raw data into data objects that are meaningful to the researchers. In addition to that, the streams framework also provides some inherent, predefined attributes of these data objects. For example, the streams framework identifies a *trip* as a sequence of individual geolocation data points that has a meaningful start and end point (what we call *touchpoints*). From this trip, the streams framework automatically calculates attributes such as its distance and duration. This information is then sent back to the central data platform and stored such that the researcher
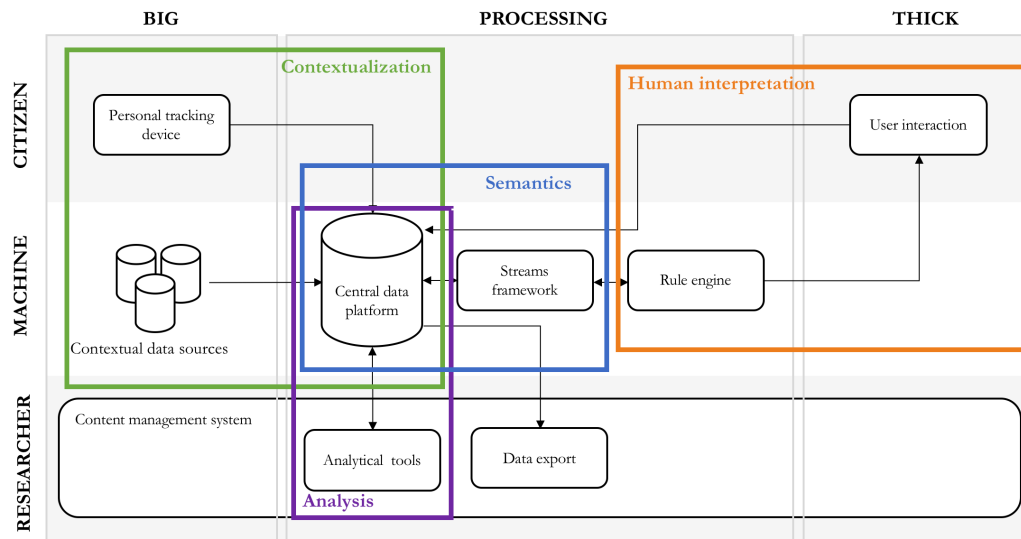
Figure 4. Representation of how the Citizen Toolbox allows for sensemaking in four ways.
© Authors' illustration.

can work with these meaningful objects and their attributes. In our case example, we could identify the school and a student's home as *touchpoints* and consequently identify one's *trip* from home to school.

Analysis - Analyzing the available data in the central data platform is the third sensemaking strategy supported by the Citizen Toolbox. Depending on the aim of the researcher, data analytics could be used either as exploratory or confirmatory. The former tries to find patterns and relations in the data, whereas the latter relies on statistical techniques to validate a hypothesis. An important aspect here, is that the results of the analysis can be added to the existing data. For example, when we segment the students based on the average duration of their trip to school, we can add this segmentation information to their profiles.

Human interpretation - Finally, we come to the real human sensemaking, allowing to add a human interpretation layer to the data consisting of experiences, emotion, motivations, etc. The rule engine allows to act upon all the previously identified data objects with their corresponding attributes by means of rules that trigger an action. For example, when we want to investigate if roads with heavy traffic are perceived as more unsafe, we could define the following rule: if (traffic = heavy) and (transport_mode = bike) and (location = school) then send_question. This means that if the condition is satisfied, the student will receive a question through the application installed on his or her smartphone upon arrival at school: "Did you feel unsafe on the road today?". The student can answer this question by simply choosing for yes or no and send the response.

We hope that it has become clear to the reader that these four sensemaking strategies in the Citizen Toolbox are not independent of one another and the result of one could be the

input for another one. In this way, the thickening process becomes a cumulative one where the big data gradually becomes thicker.

CONCLUSIONS AND THOUGHTS

In this paper we described the Citizen Toolbox as a research tool to combine big and thick data as to enable human sensemaking. This toolbox is a supporting tool for user researchers and allows them to collect objective, big data and use this data to interact with the end-users to make sense of this data by enriching it with thick data. This empowers researchers by providing them a better understanding of users' experiences and allowing them to act upon the insights resulting from the integration of big and thick data. The additional value generated by this integration has been discussed in the field for a while and the benefits are clear: large-scale in-depth insights. Thick data allows for an in-depth understanding of people's beliefs and behaviors in their context, while big data allows to detect patterns in large sets of data points.

The core strength of this tool is its ability to combine different data sources and automatically act upon them by means of the rule engine and the interaction module. The tool allows these interactions to be triggered within a specific context, which empowers the researcher in collecting the thick data. The Citizen Toolbox hereby supports researchers to cope with the asynchronicity problem that we have identified in several Urban Living Lab projects. Moreover, the toolbox allows for four types of sensemaking: contextualization, semantics, analysis and human interpretation. These strategies interact with each other and thereby create an additional value in terms of understanding. In other words, these strategies allow to thicken the big data. However, the question is how thick did it become? How can we assess this and when do we reach the limits of the big data feeding the thickening process and do we eventually need to turn to other methods anyway? These are still open questions to us.

We developed the Citizen Toolbox with the goal of being a supportive tool for research, which might have resulted in less attention to the actual research approach. However, we believe that the configuration of the toolbox allows for an implementation to support multiple research approaches in various types of research projects. It is up to the researcher to decide how and when the Citizen Toolbox can be used. From our point of view, we consider the use of the Citizen Toolbox to be relevant to gain more in-depth insights from end-users in end-product development and evaluation, behavior mapping and understanding and the evaluation of behavior change interventions. Moreover, the modularity of the data input allows the Citizen Toolbox to be used in multiple research domains, such as for example the health domain where geolocation might not be the most important personal data input, but rather physiological data such as heart rate or galvanic skin response.

We want to end this paper with two final matters of concern related to ethnographic research: the level of serendipity and the balance between privacy and intrusiveness. First of all, the Citizen Toolbox is designed to be serendipity enabled, because it allows to induce additional insights from the data and act upon it in a later stage. However, at the same time, the set-up of the toolbox does require some initial scoping and pre-definition of rules and variables. This does not only threaten serendipity, it also risks to cause single-mindedness by the researcher to only look for confirmation of the hypotheses. To overcome this, we value

the combination of the use of the Citizen Toolkit with other qualitative research methods, such as focus groups or in-depth individual interviews.

Finally, we have to take into account the balance between privacy and intrusiveness. On the one hand, the continuous stream of data coming from the personal tracking device, allows us to use a less intrusive approach, because we can monitor a lot of information without having to ask the participant everything we are interested in. However, on the other hand, this raises some justified questions regarding participants' privacy. Although the toolbox includes measures to protect participants' privacy, it remains a sensitive matter and potentially more complicated, because of new EU regulations. The question is where we have to keep the balance and if it is independent of the research project. For us, this remains an open question for the time being and finding this balance will continue to be a challenge in many research projects involving human subjects and their personal data.

Annelien Smets studies the relationship between society and data-driven innovations. She holds a master's degree in Information Management and Artificial Intelligence from KU Leuven and currently works at imec-SMIT, Vrije Universiteit Brussel as a smart city researcher.

Bram Lievens is a senior user researcher at imec-SMIT, Vrije Universiteit Brussel since 2002. He is involved in various research projects investigating the interplay between technology, society and humans, from a user-centered perspective. He has a bachelor's degree in social and cultural work and a master's degree in communication sciences.

NOTES

1. More information on the City of Things project in Antwerp (Belgium) can be found online at https://www.imec-int.com/en/cityofthings.

REFERENCES CITED

Batty, Michael
2016   Big data and the city. UK: Alexandrine Press.

Blommaert, Jan and Jie Dong
2010   Ethnographic fieldwork: A beginner's guide. UK: Multilingual Matters.

Boonen Michelle and Lievens Bram
2018   The use of live-prototypes as proxy technology in smart city living lab pilots. *Human-computer interaction international conference* 2018.

Churchill, Elisabeth
2017   The Ethnographic Lens: Perspectives and Opportunities for New Data Dialects. EPIC Perspectives. https://www.epicpeople.org/ethnographic-lens

Coenen, Tanguy, van der Graaf Shenja and Walravens Nils
2014   Firing up the city-a smart city living lab methodology. *Interdisciplinary Studies Journal* 3(4), 118.

Curran, John
2013   Big data or 'big ethnographic data'? Positioning big data within the ethnographic space. *Ethnographic praxis in industry conference proceedings* 2013, 62-73.

de Oliveira, Álvaro, Campolargo Margarida and Martins Maria
2015   Constructing human smart cities. *Smart Cities, Green Technologies, and Intelligent Transport Systems,* 32-49.

Dervin, Brenda
1992   From the mind's eye of the 'user': The sense-making qualitative-quantitative methodology. *Qualitative research in information management,* 61-84. Englewood, CO: Libraries Unlimited.

Du, Jia Tina
2014   Library and Information Science Research in Asia-Oceania: Theory and Practice. IGI Global.

Hey, Tony, Tansley Stewart and Tolle Kristin M.
2009   The fourth paradigm: data-intensive scientific discovery. Redmond, WA:Microsoft research.

Juujärvi, Soile and Pesso Kaija
2013   Actor roles in an urban living lab: What can we learn from Suurpelto, Finland*?. Technology Innovation Management Review* 3(11), 22-27. doi: 10.22215/timreview742

Kari, Jarkko
1998   Making sense of sense-making: From metatheory to substantive theory in the context of paranormal information seeking. *Nordis-Net workshop: (Meta) theoretical stands in studying library and information institutions: individual, organizational and societal aspects.* Oslo, Norway.

Kontokosta, Constantine E.
2016   The quantified community and neighborhood labs: A framework for computational urban science and civic technology innovation. *Journal of Urban Technology* 23(4), 67-84. doi: 10.1080/10630732.2016.1177260

Latré, Steven, Leroux Philip, Coenen Tanguy, Braem Bart, Ballon Pieter and Demeester Piet
2016   City of Things: an integrated and multitechnology testbed for IoT smart city experiment. *Smart Cities Conference (ISC2), IEEE International*, 1-8. doi: 10.1109/ISC2.2016.7580875

Massachusetts Institute of Technology
2008   Context-aware experience sampling. http://web.mit.edu/caesproject/index.htm

Pierson, Jo and Lievens Bram
2005   Configuring living labs for a 'thick' understanding of innovation. *Ethnographic Praxis in Industry Conference Proceedings* 2005, 114-127

Ratto, Matt
2000   Producing users, using producers. *Participatory Design Conference New York* 28(11).

Rivoal, Isabelle and Salazar Noel B.
2013   Contemporary ethnographic practice and the value of serendipity. *Social Anthropology* 21(2), 178-185.

Smets, Annelien, Lievens Bram and D'Hauwers Ruben
2018   Context-aware experience sampling method to understand human behavior in a smart city: a case study. *Measuring Behavior 2018 Conference Proceedings*, 323-329.

Spagnoli, Francesca, van der Graaf Shenja and Ballon Pieter
2017   Changing behavior in the digital era: a concrete use case from the health domain. *World Academy of Science, Engineering and Technology International Journal of Psychological and Behavioral Sciences* 11.

Veeckman, Carina and van der Graaf Shenja
2014   The city as living labortory: A playground for the innovative development of smart city applications. *Engineering, Technology and Innovation (ICE),* 1-10. doi: 10.1109/ICE.2014.6871621

Wang, Tricia
2016   Why Big Data needs Thick Data. Ethnography Matters. https://medium.com/ethnography-matters/why-big-data-needs-thick-data-b4b3e75e3d7