Performing smart cities research based on existing datasets: a methodology framework

Montero, Eladio ; Lievens, Bram; Heyman, Rob; Ballon, Pieter

Published in:
IEEE International Smart Cities Conference

DOI:
10.1109/ISC246665.2019.9071772

Publication date:
2019

Document Version:
Accepted author manuscript

Performing smart cities research based on existing datasets: a methodology framework

1st Eladio Montero Porras
imec-SMIT-VUB
Vrije Universiteit Brussel
Brussels, Belgium
eladio.montero@vub.be

2nd Bram Lievens
imec-SMIT-VUB
Vrije Universiteit Brussel
Brussels, Belgium
bram.lievens@imec.be

3rd Rob Heyman
imec-SMIT-VUB
Vrije Universiteit Brussel
Brussels, Belgium
rob.heyman@vub.be

4th Pieter Ballon
imec-SMIT-VUB
Vrije Universiteit Brussel
Brussels, Belgium
pieter.ballon@vub.be

***Abstract*—Our understanding of how a city works is changing as more and more activities and interactions of citizens with and within the city are being tracked. It is estimated that by 2020, 1.7MB of data will be created every second for each person [1]. These data are created by different actors, including businesses that collect information to fuel their services. There is a possibility for cities to enrich the understanding of their citizen dynamics and improve data-driven decision-making using these data. However, cities do not have the resources or skills (yet) to handle and analyze this new type of information. In this paper, we propose a methodology for cities to map and identify different sources of existing data, in order to give them a new meaning in the urban planning context. Also, we present our empirical experience with the methodology in solving issues at the city management level. Results show that cities can not only gather new insights by exploring existing datasets from businesses but also give them a new purpose to support and evaluate decision-making on the urban space and its development.**

***Index Terms*—smart city, data-driven policy making, data re-usability, data-analytics**

I. Introduction

The study of human behavior in a city, and especially the analysis of patterns of how citizens coexist with their surroundings and social context has gained more and more interest in the past years, both from (social) science as from city administration perspective.

Traditional quantitative methods to evaluate these activities such as surveys, questionnaires, and censuses have served well to research human interactions. These methods have some limitations, especially with the size of the sample, missing data issues [2], and these methods only allow to collect information at a certain point in time with several weeks or months of lag [3]. These limitations cause difficulties to study complex human behavior since they can be susceptible to biases, like recall bias that might affect the accuracy of the outcomes [4].

Contrary to these methods, cities have adopted the use of new technologies to collect information to support their decision-making. In what nowadays is called the smart city, different devices, such as sensors, surveillance cameras, automatic street lighting and many other embedded electronics (often under the umbrella term of Internet of Things) are being implemented to monitor and track the activity of the activities happening in the physical and digital space [5]. According to Gartner, it is expected that for 2020, there will be 20.4 billion connected devices [6].

These connected devices produce an enormous set of (often real-time) data. They are essential to enable big data analytics to extract insights and detect trends of behavior over time [7]. For public organizations, these analytics can increase transparency in the operations, and improve processes by adding more knowledge and to make more informed decisions and optimize resources [8].

However, despite its promise, it has been reported that currently less than 50% of this so-called structured data (relational databases, tables in text files, spreadsheets) is being used in decision-making processes and less than 1% of the unstructured data (e.g. text, video including surveillance, audio, mobile signals or social media posts) is being used at all [9].

There is, however, an alternative between using traditional quantitative methods and investing in gathering data from a plethora of devices. In this paper, we review the question: how can cities make use of available data from businesses for decision-making and research? We propose a general methodology to explore the use of commercial and structured data for smart cities applications, decision-making in cities and social science research in urban studies. To do so, we will review a use case for the use of data from a company to assess the impact of an event for the city Ghent, Belgium in 2018.

The next sections of this paper will explore the early work done by other works, followed by describing the proposed methodology and the empirical results of the implementation. Lastly, it will conclude with some implications and lessons learned, and the challenges for future implementations.

A. Early commercial data use for city management

We define commercial data as the data that is already being collected by different businesses to fuel their services and applications. For example, mobile operators collect data on calls, text messages and the timestamp of these events whereas credit cards companies collect data on purchases, consumers, shops and other information crucial to run their business.

Taking advantage of different sources of available commercial data for social research has been a practice widely

used in different works and fields, for example in economics [3], [10] and [11] used data from Google Searches to predict and study economic issues like unemployment, automobile demand, inflation and vacation destinations. Cities also use data from companies like Twitter also offer their data to analyze phenomena in areas such as law enforcement, tourism, and politics [12].

The tremendous increase in mobile cellphone use in the last decade, for example, enables the visualization of mobility patterns and plan better transportation strategies. It is also proven possible to analyze data from mobile providers to detect human activity places and mobility patterns [13] and [14].

In the deployment of IoT and (city-based) services we do see that companies are the primary driver towards smart cities and its technological foundations. One of the most well known and controversial cases is Google Sidewalk Labs, where they use the intelligence gathered by Google and other sensors to generate insights on a large-scale urban experiment in the Waterfront district in Toronto, Canada [15].

There is an increasing multitude of companies, small and big, that are entering this area by providing an enriched datastream. For example data from MasterCard has been used to study the demand management of public transportation in Chicago, and identify how commuters behave and make choices [16]. A similar scenario happened in Cincinnati with data from Uber to improve road safety [17] and traffic congestion in Washington DC [18]. Strava, a mobile application for cyclists and runners to track their routes, helped over 100 transportation planning departments around the world who are using their platform for urban planning [19].

These cases illustrate that there is already a lot of useful data available, but that they need to be looked at and analyzed from a different perspective.

Using available data from businesses certainly has benefits over other sources of data. Since many organizations quantify the interactions of their businesses with their consumers, there are many possibilities for analysis at a minimal cost. Wang, Calabrese, Lorenzo & Ratti [20] noted that as these data are already being produced for their core business, it is relatively cheap for re-use. Also, when done with proper privacy compliance, these uses do not affect either the normal business nor the consumer. These type of (commercial) data in general, can fill gaps where traditional methods cannot reach, and the amount of data available produced by different actors already makes almost every aspect of human life quantifiable [21].

II. Methodology

In this paper, we propose a methodology for city administrators to take advantage of structured data from businesses to look for opportunities and fuel their research and decision-making strategies. The steps of the methodology are summarised in Figure 1.

Our methodology assumes that data is structured, this means that the database is stored in a rigid place and it is divided by fixed fields and columns and it can be queried for its provenance and meaning. While these requirements can be seen as limits, we believe that these are necessary to safeguard the quality of the data and its output. However, a preliminary process of data cleaning always needs to be executed.

We also require support from the vendor or provider of data. This means that documents like data dictionaries, database architectures and glossaries are available. Ideally, there is an effective communication channel with the data provider, so the researchers can validate their assumptions (take into account that these data are given a new meaning) and communicate timely the business rules that apply to the data.

Another aspect not covered in this methodology is the way city administrators and companies get to collaborate together, or how they reach that agreement. Those administrative issues are beyond the scope of this paper, here it is assumed that the city administrators have access to data from such companies.

Lastly, we assume that the data provided has been vetted to be compliant with the GDPR in case it contains personal data[1].

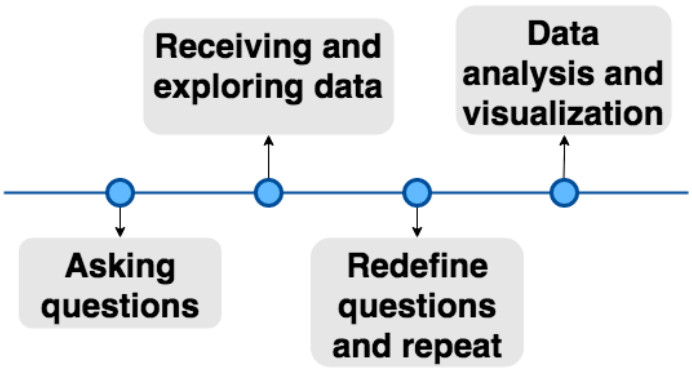

Fig. 1. Four steps of the methodology re-use data.

A. Asking questions.

Questions occupy a central and evolving role in any decision-making process. In this and other methodologies for quantitative and qualitative research, questions serve as a form to give direction and focus to a study [23]. At the same time, it sets boundaries to the type of data that needs to be collected [24].

The question identified should entail a clear problem of interest to a certain audience [25]. They should be derived from a plausible framework and therefore discussions with experts and detailed research required. In addition, it is important that the defined questions or not to open and thus answerable. If not, it will not be possible to provide the proper directions on what to look for in the data, selecting the proper data and choose the appropriate analysis methods.

[1]In the light of the GDPR legislation, there are already guidelines for data sharing between businesses and governments (B2G). As they note: business-to-government data collaboration agreements should seek to be mutually beneficial while acknowledging the public interest goal by giving the public sector body preferential treatment over other customers [22]

Leek & Peng [26] argue that the most common reason why most data analyses fail is because of questions poorly defined or mistakenly chosen. Therefore it is necessary to identify specific (research) questions, which can be refined into concrete sub-questions. This will further aid in evaluating whether the identified goals and requirements can be met. Starting with the most general questions and goals, the goal is to narrow down how they can be measured. One big idea or goal can be composed of several sub-questions, that will shape the future interpretation of the data analysis.

B. Receiving and exploring data

In this step, it is necessary to collect data. Companies usually provide an API (Application Programming Interface) as a communication protocol for the structured data to go from the company to third parties. Other possibilities are database snapshots or text files with a subsection of the data. The most cost-effective transfer of data for businesses is to give it as it is. Therefore most companies do not format their data for other purposes than their own.

But regardless of the format that the company agrees to share the data, this should be flexible enough to allow addressing many different questions and test multiple hypotheses. It is not useful, for example, if the dataset only allows answering one specific question, and not have the opportunity to dig deeper into the preliminary results.

Exploring the data received allows to think about what meaning can be gained out, what questions could be answered and to assess them, also what modifications in the data are necessary to do so. This can be done by exploring the different entities[2] in the database.

Entities can reveal information about how each business decides to describe an object or concept in their database. For example, the entity of person can be described in Twitter as someone that has a user handle, a number of followers and following, has a location, a list of tweets, etc. This is a representation of a person in real life, and it can be different from a database from a bank, for example. There is an opportunity to study their behavior and interactions within the city for both, just from a different perspective.

When analyzing the data, one should be clear about the rules and business-specific concepts that the data represents, so the research outputs show the correct results based on the data collected and the assumptions made. This is where data dictionaries, glossaries and a clear communication channel with the data provider come useful.

If these business rules are clear and the entities of interest have been identified, the second step that is required is to check how clean and well structured the data is. In many cases, the data provider has to notify how the data should be prepared for its first-time use. Also, in this part, a validation if an entity of interest is correctly represented in the database needs to be performed. This will depend heavily on the data provider core business.

[2]Entity is a concept from database design that describes a representation of a concept or object of the real world in the database, and their information about it that is relevant for the system [27]. For example, for credit card data, some entities can be client, card or transaction.

C. Redefine questions and repeat

It should be noted that all previous steps can be seen as a feasibility check; will the data be able to deliver an acceptable answer? Next steps consist of further defining the questions so that these can be operationalized in such a way that they can be transformed into specific data-analysis processes.

The questions themselves function as a back and forth object between different parties with different background knowledge. The defined questions are designed as a boundary object [28]. A boundary object is an object that is flexible enough to be understood and worked on by all actors despite their different backgrounds. Secondly, it also allows more autonomous actors to further drill down specific parts that relate to the commonly defined goals. This process can be difficult, especially from the computational social science perspective. Some policy or social science concepts are too complex and broad to define in concrete terms, especially with re-used data.

This is why we propose an iterative process in order to reach an agreement about initial questions that may not be directly observable from the data. Inspired by the concepts of user stories and use cases in software development frameworks like Scrum and Extreme Programming, the objective is to help define the requirements for the data analysis. The intention is to describe the functionality in natural language, so users, decision-makers, data analysts and other stakeholders involved can agree on what is expected and how to achieve it correctly, these frameworks are created to embrace change in the development process [29].

We further define our boundary object as follows. The question format is composed of five columns (this length can change depending on the project and specific needs). It starts with defining the questions. The idea is to disassemble them to achieve more specificity as discussed in the first step. A question is answered by a set of tasks that all groups involved will help to create and agree on. The structure is presented as follows:

1) The first column is for the questions/goals. These can be general or specific. They represent the main questions that have to be derived from the analysis.
2) For each task derived from a question, they should have an identifier, so these can be traced and discussed throughout the document.
3) The name of the task so that it can be used and referred to in the discussions.
4) A basic description of the task at hand, this works as a guideline for data analysts and other groups to agree on what the analysis should contain and how should be represented.
5) A notes field for annotations, to keep track of the discussions, data-processing, important rules and variables.

For the refinement of the questions, it is important that the different groups involved discuss topics like general expecta-

tions, data quality, database architecture, and business rules. Moreover, each subtask should follow the principles of the user story definition: Independent, Negotiable, Valuable, Estimable, Small and Testable (INVEST) as proposed by Wake [30].

D. Data analysis and visualization

After an agreement has been reached, the data analysis can be executed accordingly. The outputs from the data analysis will be used to check the status of the questions listed. New questions and discussions can also arise, the boundary object is a dynamic table, meaning that permanently changes can be made. By using the object in the same fashion these iterations and updates can be documented. This way, every party involved can keep track of the process.

III. Use case implementation

The above methodology was used in an empirical experiment in which wanted to assess to what extent existing data could be used to assess the impact of a city event. The goal was to describe the impact of an event in a city, before, during and after it to see if there was a sustainable effect in their economy. To do so, we took the example of the city of Ghent in Belgium and the event was the Christmas holiday of 2018, (November 23rd to December 26th) compared with two months before and two months after.

A. Asking questions

After defining the context, we decided to create our analysis in terms of seven sub-questions and goals.

1) Is there an increase in consumers during the event?
2) Did the event attract new consumers?
3) What shops benefited from the event?
4) What areas or zones in the city were mostly visited during the event?
5) What are the profiles of the consumers?
6) Was the number of spending higher/lower?
7) Is there a sustainable effect?

It can be seen that these questions can still be more specific, and their type is mostly descriptive, to explore the interactions around the event and its implications on the data at hand.

B. Receiving and exploring data

For the data, we used existing data being collected by a loyalty program provider (e.g. Joyn) They provide a customer loyalty program for shops and consumers, and a loyalty card that is used when consumers make transactions in shops all over Belgium.

Joyn acted as our data-provider, sharing their data dictionary and a document detailing the database architecture, with all the business rules and guidelines to perform queries on their database. The company also agreed to share a snapshot of their database, personal information like email, name, home address, and telephone was taken out of the tables to guarantee the privacy of its customers.

Before any analysis, it was necessary to study the database architecture, so we could identify different entities important for the company such as shops, merchants, transactions, consumers and rewards, among others.

By looking at the entities, there was potential to study in terms of spending, what was the impact for Ghent looking from the perspective of active consumers, number of transactions and amount in euros of each transaction, active city zones and shop types. However, we found out some insights that we partially suspected, for example, the information about the consumers such as age, gender, and city of origin was either too dirty, empty, or contained seemingly fake records. Also, by looking at the business rules and in interaction with the data-provider, we realized that not every transaction was possible to infer the amount in euros spent on the purchase.

Despite this, we could find several advantages of using this data: there was plenty of records for several years, also even though the biggest retailers were not in the app, and not every person in Belgium used Joyn, there was a large number of consumers and different types of retailers.

C. Redefine questions and repeat

After deciding that there were interesting opportunities at the data, we transformed the questions into specific tasks. For this, there was a back-and-forth process that took considerable time to identify the proper questions and analysis. In total, for the analysis of the impact of the event, there were 20 specific tasks derived from the initial questions defined in step 1.

As an example, the fifth question defined in the first step (what are the profiles of the consumers?) was subdivided into five tasks that the data in hand can correctly answer. The structure defined ended as follows:

- Consumers' locality status: define if the consumer is a citizen of Ghent not.
- Number of transactions week: for local citizens and for visitors, the number of transactions done by each group per week.
- Euros spent per week: where possible, compare the amount in Euros by visitors and local citizens per week.
- The variety of shops consumers visit: compare the total number of transactions visitors and local citizens per category of the shop.
- Areas of consumption: identify which places the visitors and the local citizens visit during the Christmas period.

D. Data analysis and visualization

After the data analysis, we could address our basic question of how this Christmas period impacted the city of Ghent and more concretely the impact of visitors and shopping. Not only could we gain knowledge of the number of visitors, but also in terms of the purchases done, the type of shops as well as the places visited in the city. Due to the fact we had access to an existing dataset, we could not only look into the Christmas period itself but see how it deviated with the periods before and after and by so specifically looking at how it impacted the city. Figure 2 shows one of the maps used to visualize consumption areas by visitors in Ghent.

These insights are helpful for the city of Ghent so they can identify the effect of such period and supporting events in terms of the type of visitors, which areas are most visited and when which stores benefit the most, etc. By so they can adjust their strategies for the future. One of the follow-up changes that was derived from this analysis was to compare Ghent with other cities in Belgium and by so put the results more in perspective and make them more valid. For example, to what extent were the outcomes limited to the city of Ghent or a nation-wide trend.

Fig. 2. Heat map of the consumption areas by visitors in the city of Ghent, for the Christmas holiday of 2018.

IV. Conclusions

In this paper, we presented a methodology for cities on how to take advantage of the increasing amount of data that is being collected today and to use it for assessing f.e. policy decisions or strategies, without setting up a new data collection process. The use of a boundary object appeared to be crucial in this.

However, there are some challenges in doing so. While there are many benefits in the (re-)use existing data, and specifically data analytics, it is not always ann easily transferable good. Not only in terms of accessing this data but also in terms of the proper interpretability in available data as the method tries to use data in such a way it was not directly intended to. A good understanding and knowledge of the exact meaning of the user data are crucial.

For this reason, data exploration is the most time-consuming step of the methodology. Time needs to be invested to explore and get familiar with new data architecture and on top, giving new meaning to its original purpose. This will also require a good interaction with all parties to have a clear understanding of the data so that the correct assumptions can be made.

The boundary object also needs to bridge different worlds and be able to align the different parties involved. As, especially when applied in a city administration context, people with different expertise and skills will need to have a common understanding of the data and its meaning as well as the questions to be addressed. This requires an iterative process and interactive dialogue between all partners.

As we start from the assumption of the potential use of existing data one needs to accept that it will not be able to address or answer all questions identified. Data will have its limitations and trade-offs will have to be made. For example, not every set of data is able to yield results that are applicable to the general population, or they might be too biased to interpret correctly.

As they approach re-use data, compliance data privacy legislation (in this case the GDPR) is needed. Especially the GDPR principle of purpose compatibility between old and new use is something that needs full attention and requires specific skills to investigate.

For future work, it is needed to further elaborate on the methodology and test its adaptability to different scenarios and solving a wider range of problems. Points of attention here are 1. ways of structuring the data and assessing the through the meaning of data, 2. further built upon a common understanding on the specific desired type of data-outcomes and 3. create an assessment from a legal standpoint, to assess the compliance with current (privacy) rules and regulations.

Acknowledgment

This research is part of the Smart City Chair program of IMEC-SMIT hosted at the Vrije Universiteit Brussel (www.smartcitychair.be) and supported by Joyn, one of its sponsors.

References

[1] DOMO, "Data Never Sleeps 6 | Domo." [Online]. Available: https://www.domo.com/learn/data-never-sleeps-6

[2] L. Einav and J. Levin, "Economics in the age of big data," *Science*, vol. 346, no. 6210, p. 1243089, Nov. 2014. [Online]. Available: https://science.sciencemag.org/content/346/6210/1243089

[3] H. Choi and H. Varian, "Predicting the Present with Google Trends," *Economic Record*, vol. 88, no. s1, pp. 2–9, 2012. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1475-4932.2012.00809.x

[4] A. Smets, B. Lievens, and R. D'Hauwers, *Context-Aware Experience Sampling Method to Understand Human Behavior in a Smart City: a Case Study*, Jun. 2018.

[5] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, "Internet of Things for Smart Cities," *IEEE Internet of Things Journal*, vol. 1, no. 1, pp. 22–32, Feb. 2014.

[6] R. van der Meulen, "Gartner Says 8.4 Billion Connected "Things" Will Be in Use in 2017, Up 31 Percent From 2016," Jul. 2017. [Online]. Available: https://www.gartner.com/en/newsroom/press-releases/2017-02-07-gartner-says-8-billion-connected-things-will-be-in-use-in-2017-up-31-percent-from-2016

[7] S. Jeble, S. Kumari, and Y. Patil, *Role of Big Data in Decision Making*, Jan. 2018, vol. 11.

[8] E. Al Nuaimi, H. Al Neyadi, N. Mohamed, and J. Al-Jaroodi, "Applications of big data to smart cities," *Journal of Internet Services and Applications*, vol. 6, no. 1, p. 25, Dec. 2015. [Online]. Available: https://doi.org/10.1186/s13174-015-0041-5

[9] L. DalleMule and T. H. Davenport, "Whats Your Data Strategy?" *Harvard Business Review*, no. MayJune 2017, May 2017. [Online]. Available: https://hbr.org/2017/05/whats-your-data-strategy

[10] N. Askitas and K. F. Zimmermann, "Google Econometrics and Unemployment Forecasting," Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 1415585, Jun. 2009. [Online]. Available: https://papers.ssrn.com/abstract=1415585

[11] G. Guzman, "Internet Search Behavior as an Economic Forecasting Tool: The Case of Inflation Expectations," Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 2004598, Nov. 2011. [Online]. Available: https://papers.ssrn.com/abstract=2004598

[12] A. A. Nuaimi, A. A. Shamsi, A. Shamsi, and E. Badidi, "Social Media Analytics for Sentiment Analysis and Event Detection in Smart Cities," 2018.

[13] J. Steenbruggen, E. Tranos, and P. Nijkamp, "Data from mobile phone operators: A tool for smarter cities?" *Telecommunications Policy*, vol. 39, no. 3, pp. 335–346, May 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0308596114000603

[14] S. Jiang, J. Ferreira, and M. C. Gonzalez, "Activity-Based Human Mobility Patterns Inferred from Mobile Phone Data: A Case Study of Singapore," *IEEE Transactions on Big Data*, vol. 3, no. 2, pp. 208–219, Jun. 2017.

[15] Sidewalk Labs, "Sidewalk Labs," 2019. [Online]. Available: https://www.sidewalklabs.com/

[16] MasterCard Inc., "Smart Cities Solutions & Technologies by Mastercard," 2017. [Online]. Available: https://www.mastercard.us/en-us/about-mastercard/what-we-do/global-smart-cities.html

[17] Uber Movement, "Improving Road Safety in Cincinnatis Northside neighborhood," May 2019. [Online]. Available: https://medium.com/uber-movement/oki-8b77c95bb368

[18] ——, "The effects of DC Metrorail service disruptions on traffic congestion," Jan. 2017. [Online]. Available: https://medium.com/uber-movement/the-effects-of-dc-metrorail-service-disruptions-on-traffic-congestion-8a14c8d5fa7c

[19] Strava Inc., "Strava Metro," 2018. [Online]. Available: https://metro.strava.com/

[20] H. Wang, F. Calabrese, G. D. Lorenzo, and C. Ratti, "Transportation mode inference from anonymized and aggregated mobile phone call detail records," in *13th International IEEE Conference on Intelligent Transportation Systems*, Sep. 2010, pp. 318–323.

[21] D. Arribas-Bel, "Accidental, open and everywhere: Emerging data sources for the understanding of cities," *Applied Geography*, vol. 49, pp. 45–53, May 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0143622813002178

[22] European Commission, "COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS. COM(2018) 232 final," Apr. 2018. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52018DC0232from=EN

[23] P. R. Lowenthal and N. L. Leech, "Mixed Research and Online Learning: Strategies for Improvement," *Online Education and Adult Learning: New Frontiers for Teaching Practices*, pp. 202–211, 2010. [Online]. Available: https://www.igi-global.com/chapter/mixed-research-online-learning/36888

[24] A. Onwuegbuzie and N. Leech, "Linking Research Questions to Mixed Methods Data Analysis Procedures 1," *The Qualitative Report*, vol. 11, no. 3, pp. 474–498, Sep. 2006. [Online]. Available: https://nsuworks.nova.edu/tqr/vol11/iss3/3

[25] R. D. Peng and E. Matsui, *The Art of Data Science*. lulu.com, Jul. 2016.

[26] J. T. Leek and R. D. Peng, "What is the question?" *Science*, vol. 347, no. 6228, pp. 1314–1315, Mar. 2015. [Online]. Available: https://science.sciencemag.org/content/347/6228/1314

[27] IBM, *IBM Dictionary of Computing*, 10th ed. New York, NY, USA: McGraw-Hill, Inc., 1993.

[28] S. L. Star, "This is Not a Boundary Object: Reflections on the Origin of a Concept," *Science, Technology, & Human Values*, vol. 35, no. 5, pp. 601–617, 2010. [Online]. Available: https://www.jstor.org/stable/25746386

[29] K. Beck, "Embracing change with extreme programming," *Computer*, vol. 32, no. 10, pp. 70–77, Oct. 1999.

[30] B. Wake, "INVEST in Good Stories, and SMART Tasks - XP123," Aug. 2003. [Online]. Available: https://xp123.com/articles/invest-in-good-stories-and-smart-tasks/