



# ARTIFICIAL INTELLIGENCE USAGE OPPORTUNITIES IN SMART CITY DATA MANAGEMENT

Luca F. Hudasi and László Ady\*

NextTechnologies Ltd. Complex Systems Research Institute  
Budapest, Hungary

DOI: 10.7906/indecs.18.3.8  
Regular article

*Received:* 17 February 2020.  
*Accepted:* 28 June 2020.

## ABSTRACT

In our current study, we are aiming to explore data management methods in Smart City systems. In data management, Artificial Intelligence can be used as well. Solutions for the usage of Artificial Intelligence and integration into Smart City concept will be researched as well. Main motivation of the study is to draw attention to one of the most important element of Smart Cities, to the flow of data. Our study provides a possible solution for managing data and keep data up-to-date in such systems with the usage of newest technology possibilities. While explaining the solution, we will give all the necessary details about the data flow model between the Artificial Intelligence based system and humans who are using the Smart City. For managing the dataflow, we would like to use Big Data methods extended with other required methods. We are using the term of Big Data as a technology maximizing computation power and algorithmic accuracy to gather, analyse, link, and compare large data sets [1] connecting with Artificial Intelligence solutions.

## KEY WORDS

smart city, artificial intelligence, data, management, innovation

## CLASSIFICATION

ACM: 10010147.10010178

JEL: C8

\*Corresponding author, *η*: [adylaszlo@nexttechnologies.hu](mailto:adylaszlo@nexttechnologies.hu); -;  
H- 2234 Maglód, Sugár út 44.

## INTRODUCTION

Intelligent systems, such as Smart Cities are based on the flow of information [2]. Another important aspect of a good smart city is to make good decisions. It is logical, that if a system has bad or missing information, it cannot make good decisions [3].

‘The first step in a city becoming a “smart city” is collecting more and better data.’ [4] says John Walker in his study. Therefore the following main areas are important to cover upon collecting data:

Develop an automata data collector system.

Develop a people triggered data collector system.

Develop a data sharing and correction system.

The following figure shows the flow of data based on collection and sharing:

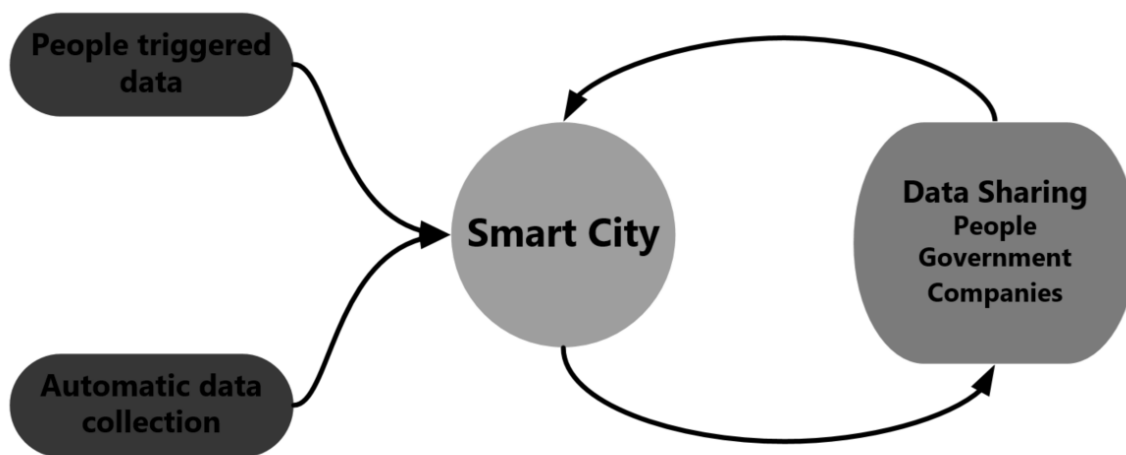


Figure 1. Data flow of a smart city.

On figure 1, arrows show the direction of data flow. Although people triggered data and Automatic data collection is mainly one directional, data sharing is not. Data Sharing part has many subcomponent and the data flow in this case bidirectional since the participants not just providing but getting data as well. In the following sections we discuss the details of each components.

## DEVELOP AN AUTOMATA DATA COLLECTOR SYSTEM

Developing an automata data collector system covers the already well-known methods, such as having traffic monitoring systems, automata government administration bodies (for voting, for taxes, etc.), and an automata traffic-, weather forecast-, and energy distribution system. To build fully automated systems using big data, it is a requirement to have a built-out sensor network such as camera, temperature measurers, motion detectors and GPS based devices [5]. It does not mean these systems do not require a supervisor, but they can operate in an independent way and processing data coming from sensors. Many of these systems are already existing in testing or in live format, for example in Singapore, London, Barcelona, etc. [4, 6]. It is important to use the experiences from these cities to build a more reliable one.

## DEVELOP A PEOPLE TRIGGERED DATA COLLECTOR SYSTEM

People triggered data collectors are Intelligent Systems where the information is coming from an active human input. Such as if an inhabitant visit an authorized system (through internet) and report or handle an administrative hacc they need. For example request a credit, rent a

car, report a misdemeanour activity, etc. Here, the smart system has to be prepared to be able to serve the customer's need and make fair and clear decisions. This is the most dangerous form of Data Collection because the final decision is made by an Artificial Intelligence (AI) using a learnt scheme. The discrimination factor is too high, and there are already case studies for detecting and removing discriminative part from the software [7]. 'Some companies curtailed their customers' credit if charges appeared for counselling, because depression and marital strife were signs of potential job loss or expensive litigation' [8] says Racher O'Dwyer and this was just one example from the many. The question which should be considered is: is it ethical and legal to allow such kind of discrimination? If the answer is no, then the information system should be prepared to prevent this, or supervised for such actions. Connecting psychology and big data in the field of allowance is a new immature area which needs further studies and testing before essential decisions are made based on the result. Therefore, we strongly recommend to use a bipolar system in this case: first, the AI based system make the decisions using the source from collected Big Data, then a human supervisor should overview the output with the factors used in the decision, and validate or decline it. Using this 'two steps verification' looks longer, but it is not so much. Collecting data would be still the responsibility of the automata system and this is the most time-wasting part of the process. Educated human supervising would correct and develop the AI to make better decisions in the future. Once the system works measurably stable and ethical, the supervising work can be decreased.

Another important point of data collection is to collect quality data [4], otherwise the information the system's decisions are based on is corrupt or missing, therefore the decisions will be similarly wrong. To achieve this, our proposed solution is to include the inhabitants of the city to clarify data.

One area where inhabitants can participate is the social-, public administration improvements. The system would be capable of filling out data and do pre-tasks for the inhabitants (for example: doing the tax, requesting for new social cards when the existing ones are going out-of-date, providing public utility usages, making renting, other billing tasks, etc.) but the citizen would have the opportunity to monitor the decisions, and correct them if needed. Next time the system would learn from the mistakes and from the habits of the people, and would make better decisions.

Another are would be for extra comfort services, where people would voluntarily provide information for the system which then can help them to take away tasks from their shoulders, such as organizing trips, ordering and delivering food, other supplies, or appointments with doctors or similar. With more up-to-date corrections of the information and decisions the system is operating with are made, the better intelligent services could be provided. It is always very important to leave an opportunity to supervise the decisions the AI is doing in the place of people, to avoid discrimination and bad decisions (detailed above).

## **DEVELOP A DATA SHARING AND CORRECTION SYSTEM**

To be able to cooperate with citizens in the development of data and decision-making, it is important to make the information – the smart city collects – as transparent as possible. People should see the base information of some bad decisions to be able to help to correct them.

While the smart city would provide transparent decision making for the citizens, it is critical to guard the sensitive data. Big data collection is always a hazard factor. More information the system provide, more value it represents and it will be more interesting for non-ethical parties. Therefore, any data which can be provided to third party, should be depersonalized carefully. To manage the proper depersonalization, is a key factor of the data flow. Creation of standards document for depersonalization of Smart City data is a requirement.

The main industry, which is collecting and using data is the advertisement industry. It can also produce a great income for the city by using the data collected, but the depersonalization should be act in the process here as well. People should be informed that the collected data are provided and generating income, but the system should also keep people's trust while do so. The first aim of a smart city is always to provide a better and easier life for people, get rid of discrimination and unethical decisions. Ethical and correct advertising is the part of this, but it should not lead to people exposed to direct marketing harassments which can lead to people leaving social media platforms and losing trust [9, 10].

There is another part of the industry, which is dealing with data: A company can provide other useful data for the smart city and in exchange get data as well which is important for the company. This can be a government level company as well which is collecting data and provide the result as a service for the people.

## STRATEGY TO PROCESS DATA

After discussing the aspects of collecting and managing quality data, the next important step is to have the intelligent system make good decisions. We emphasized the need for supervising decisions till they become trustable in the aspects of avoiding discrimination and improvements. We offered the idea to include citizens in the supervising work as well. Also mentioned to use up data and experience collected from other smart cities. These ideas can help maintaining data decisions. Although we have to know that a good system beside having the correct data and algorithms also needs to process data quick and effectively. A good strategy to process data is taking the following steps: data filtering, pre-processing, processing and decision support. Figure 2 shows the processing stages with their dependencies.

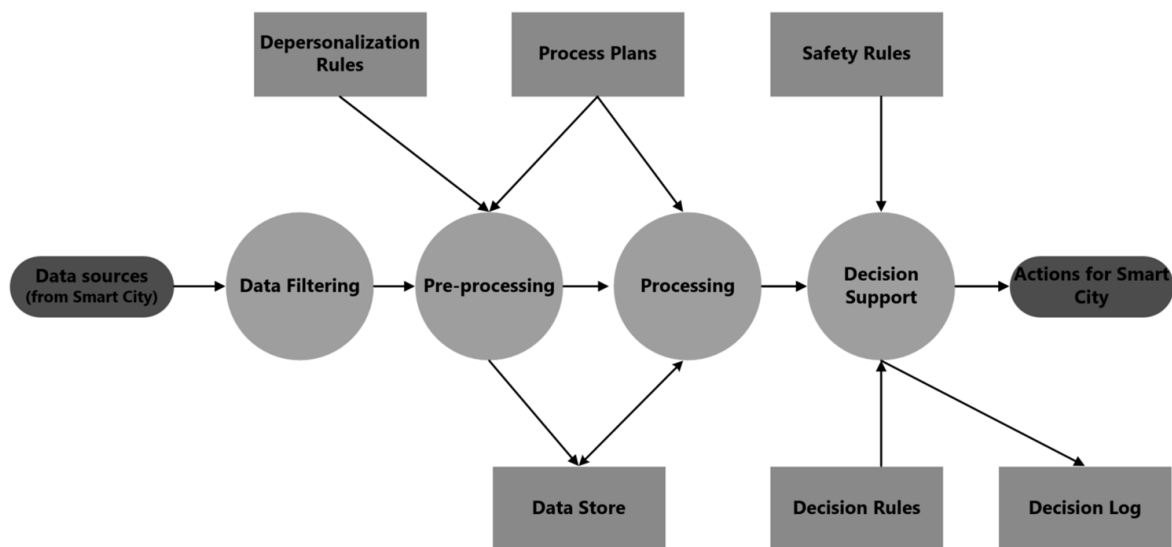


Figure 2. Logical architecture of data processing steps.

## DATA FILTERING

The aim of data filtering is to collect data which the system really needs during the processing, to minimize the amount to store and process. An existing solution to filter out too many data (mainly when the system has limitations of processing quickly) is CERN's solution. CERN had a problem of bypassing all the measured data to their servers. With Intel's help, they developed a specific FPGA which help to filter out the relevant data before bypassing it to the servers. This way, they minimized the load on the centre [11].

## **PRE-PROCESSING**

It is a good speeding strategy to execute pre-processing tasks as close at the point of data collection as possible. This means that the stage when the data comes in the system, stored procedures executed immediately and decode the input such as voice to text, picture to sort and identify, etc. Sorting (mainly of pictures) is an effective way because dedicated devices are already a stage of the sorting. As an example, a parking lot could have a dedicated camera to recognize empty spaces and another dedicated one for recognizing unauthorized parking. For pre-processing the pictures, the area reservation database needs to be presented at the moment the picture is caught and pre-processed by the camera system. This is the stage when the data depersonalization should be performed as well in case it is needed.

## **PROCESSING**

Processing has many types how it can be performed. These types can be individually used or together as well. These types are the following:

1st type: Stored procedures. In this context, stored procedure means a logically separated unit of functions for performing a specific task. The sorting component will decide which stored procedure can be executed on the data by sorting it to categories. These procedures need to be written, but on the contrary it takes less processor time while they are running on production.

2nd type: AI. During the processing, an AI module will decide which AI based stored procedure can process the data,. These AI based stored procedures will only get pre-processed data. Raw data will be pre-processed by the sorting unit. These AI units (stored procedures) are looking for relations in data.

3rd type: Data mining. Using the opportunities of data mining, the system can learn models from big data to predict problems. After detecting upcoming problems, there is a possibility to make decisions to prevent them and create a safer environment. ‘Difficulties that need to be addressed during data mining include data gathering, data labelling, data and model integration, and model evaluation’ [12]. Data gathering and data labelling can happen in the pre-processing stage, data and model integration, and model evaluation should happen in the processing stage.

4th type: Manual. Manual actions needed In cases the system can not recognize and process a certain data from the pre-processing stage (because it is not prepared for it). The system will display the details on a graphical interface and will ask a human to decide the next steps, such as sorting the data into one of the existing categories. This type of processing is more like an extend method for “error handling” together with other AI based solutions.

## **DECISION SUPPORT**

This stage of data processing is responsible to make automated decisions or help the human decision making.

Result after the processing stage is stored in a database. The decision supporting unit has the knowledge of the connections between result types and actions. Let’s take an example: There is a processing result that contains a picture that a car took a parking place. The connected action is to decrease the number of free spaces. The digital table in front of the parking lot will change and display the new data with the amount of free parking places. Or let’s take a more complex case example when the camera system detects that a big container occupied the public space near a building. The first action will be to check the permission of taking the place at the related authorities. In case there is no permission, the second action to execute is to create a draft report for the police (or related authority) and put the case up to a human supervisor to accept or decline. It is important to make this case half-automated with human

supervising, because making a punishment should not be full automated. The decision support unit has to contain a set of rules about which action can be done in which case: for example, at a heating system there could be a rule that after switching off the gas unit, it cannot be turned back in the next 5 minutes (safety period of the gas unit to chill down). If an automated decision would be made to turn on the heating because it is too cold within this 5 minutes, this rule would write it over. All the decisions which were made, should be logged for possible investigations and later improvements.

## PROTECTION OF DATA FLOW

Finally, since smart cities are operating with sensitive data, it is also a part of data management to save the data from being stolen, unauthorised modifications and destruction. Beside using the well-known defensive solutions – since the system is based on artificial intelligence and there are couple of paradigms available of normal behaviours – we can use these resources to add another level of defence for the system’s protection. We propose to build an alarm system, which – by monitoring the information flow – can alert about disharmonious data detections. Which means, if the data flowing through the system does not follow a continuously measured norm, it could be considered that the data was manipulated. This prevention method can be considered as a “software” type prevention. Physical prevention means that important data is allowed to travel on a way that is theoretically impossible to interfered or read by malicious bodies without immediate detection. In practise, this means the usage of optical cables where the network can detect any interference if the specifications of the light change. Using IEC 62443 standard is highly advised [13]. As a conclusion, we suggest to consider using the mentioned techniques and processes from this article when new smart cities are designed and build, and also for existing smart cities to develop.

## REFERENCES

- [1] Boyd, D. and Crawford, K.: *Critical questions for big data*. Information, Communication and Society **15**(5), 662-679, 2012, <http://dx.doi.org/10.1080/1369118X.2012.678878>,
- [2] Tokody, D.; Papp, J.; Iantovics, L.B. and Flammini, F.: *Complex, Resilient and Smart Systems*. In: Flammini F., eds: *Resilience of Cyber-Physical Systems. Advanced Sciences and Technologies for Security Applications*. Springer, Cham, 2019,
- [3] Tokody, D.; Schuster, G. and Papp, J.: *Smart City, Smart Infrastructure, Smart Railway*. International Conference on Applied Internet and Information Technologies, October 23, 2015. Technical faculty “Mihajlo Pupin” Zrenjanin, Zrenjanin, 2015,
- [4] Walker, J.: *Smart City Artificial Intelligence Applications and Trends*. <https://emerj.com/ai-sector-overviews/smart-city-artificial-intelligence-applications-trends>, accessed 2<sup>nd</sup> February 2019,
- [5] Neirotti, P.; De Marco, A.; Cagliano, A.C.; Mangano, G. and Scorrano, F.: *Current trends in Smart City initiatives: some stylised facts*. Cities **38**(2014), 25-36, 2014, <http://dx.doi.org/10.1016/j.cities.2013.12.010>,
- [6] Ismail, N.: *What are the most advanced smart cities in the world?* <https://www.information-age.com/advanced-smart-cities-world-123470745>, accessed 2<sup>nd</sup> February 2019,
- [7] Galhotra, S.; Brun, Y. and Meliou, A.: *Fairness Testing: Testing Software for Discrimination*. ESEC/FSE’17, 4-8 September, 2017, ACM, Paderborn, 2017,

- [8] O'Dwyer, R.: *Algorithms are making the same mistakes assessing credit scores that humans did a century ago.*  
<https://qz.com/1276781/algorithms-are-making-the-same-mistakes-assessing-credit-scores-that-humans-did-a-century-ago>, accessed 2<sup>nd</sup> February 2019,
- [9] Morrison, K.: *Consumers Don't Like and Don't Trust Digital Advertising (Infographic).*  
<https://www.adweek.com/digital/consumers-dont-like-and-dont-trust-digital-advertising-infographic>, accessed 3<sup>rd</sup> February 2019,
- [10] Tenzer, A. and Chalmers H.: *When trust falls down.*  
[https://www.ipsos.com/sites/default/files/2017-06/Ipsos\\_Connect\\_When\\_Trust\\_Falls\\_Down.pdf](https://www.ipsos.com/sites/default/files/2017-06/Ipsos_Connect_When_Trust_Falls_Down.pdf), accessed 3<sup>rd</sup> February 2019,
- [11] Barney, L.: *The FPGA based Trigger and Data Acquisition system for the CERN NA62 experiment.*  
<https://www.hpcwire.com/2017/04/14/xeon-fpga-processor-tested-at-cern>, accessed 7<sup>th</sup> February 2019,
- [12] Létourneau, S.; Famili, F. and Matwin, S.: *Data Mining to Predict Aircraft Component Replacement.*  
IEEE Intelligent Systems **14**, 59-66, 1999,  
<http://dx.doi.org/10.1109/5254.809569>,
- [13] Elder, J.: *Understanding the Importance of Physical Security for Industrial Control Systems (ICS).*  
<https://applied-risk.com/blog/understanding-importance-physical-security-industrial-control-systems-ics>, accessed 6<sup>th</sup> February 2019.