# CityPro: From Big-Data to Intelligent-Data; a Smart Approach

Mohamed Dbouk
*Faculty of Sciences - I*
*Lebanese University*
Beirut, Lebanon
mdbouk@ul.edu.lb

Mariam Hakim
*Faculty of Sciences - I*
*Lebanese University*
Beirut, Lebanon
mariamhakim181@gmail.com

Ihab Sbeity
*Faculty of Sciences - I*
*Lebanese University*
Beirut, Lebanon
isbeity@ul.edu.lb

***Abstract*— *For a city to be named smart, it has to be intelligently benefiting from the omnipresent systems available to improve life's quality and assure public safety. However, being in the world of Big Data we had to find a way that helps us in maintaining the benefits of this data while reducing its size. That's why in this paper we focus on transforming Big Data into smart data. To achieve this goal, the work of this research is dedicated to finding the optimal way of storing data in the collaborated surveillance system, CityPro. In its proposed architecture it implied using an adaptation of a federated star-schema like data repository for storage. The repository is opted to store summarized data instead of the huge amounts of data coming from partners, which will help in moving from the nomenclature of Big Data to smart data. To attain this property, we are trying to propose a new technology by creating a global unified profile for each individual in the city, which will facilitate the monitoring of suspects and predict their moves.*

***Keywords*— *smart cities, digital world, business-intelligence, big-data, smart data, smart repository, data intelligence*

## I. Introduction

Day by day, modern cities face new challenges in terms of public safety and security. Public safety is a critical feature where modern cities are exposed to emergencies like daily traffic flow, vehicle accidents, terrorist attacks, and crimes, etc. For a city with modern features, safety and peaceful living should be a priority; it must have the ability to minimize and prevent the problems that are occurring and that might occur. Every city is prone to different problems. For example, no city is accidents or crimes free; these two are popular types of threats that might interrupt public safety.

The typical surveillance techniques i.e. video surveillance incorporates thousands of cameras relying on high-speed networking infrastructures. However, they can never be smart enough to predict the happening of these problems. Therefore, if an advanced technique is adapted in a country it might be able to minimize the number of these accidents or even diminish them. This possibility creates a motive to expand and computerize surveillance techniques until a utopian technique is presented and adopted.

Almost every area in a city has its own companies and organizations. These organizations, referred to as omnipresent systems, create a network of data producers. The data produced proved to be very beneficial for the public safety of a country as stated in [1]. However, using the current systems it is only beneficial to its producer. That's why in this research we planned to begin implementing the platform called CityPro [1] by starting with the most important part which is data. The first concern in this work is where to store these huge datasets while trying to reduce their amounts and conserve their benefits at the same time. Whereas the second concern is how to integrate the data to reach a true state of benefit. Simply, we attempt to transform this Big Data into smart data.

Expanding from the naming of a modern city to a smart city can be done by intelligently benefiting from the omnipresent systems available to improve life's quality and secure a city from all its angles. It can be seen, from the perspective of technologies, as a complex arrangement of heterogeneous infrastructures and services that densely interact [2]. The ability to capture events, integrate them into an enterprise computing platform, apply fast processing, and share live real-world data presents a fundamental challenge that distinguishes alternative worlds of intelligence and smartness.

Adapting this integration requires a smart system that forms a communication bridge which not only links the systems together, but also integrates the data, reduces its amount to the most minimized beneficial form, and analyzes it to monitor and predict any upcoming problems.

These necessities nourished the idea of an integrated platform of a collaborative surveillance system, called CityPro [1]. This system is intended to protect and monitor people and public infrastructures. It is expected to:

1. Operate within live-mode by using the city digital infrastructures
2. Combine and inter-operate heterogeneous pre-existing operational systems

However, to meet these expectations a critical problem must be encountered, which is Big Data and its integration. Discussing the possibility of a solution for the Big Data integration issue arouses itself multiple problems that should be taken into consideration. These problems can be distributed into three well-known domains in the world of data, which are data security and privacy, data variety, and data volume. The main reason for such problems is the fact that the data is heterogeneous and is being collected from heterogeneous sources. CityPro is meant to be a surveillance system that requires minimum interaction with its partners [1]. This point makes it almost impossible for the actual data structure to be known, which in turn increases the level of difficulty in dealing with data integration's problem.

In this paper, Section II covers the previous and related works. We present an overview of CityPro in Section III. The work and solution we are adapting for the smart repository in section IV. Section V presents how SmartRep acts in real life with comparative scenarios. Finally, Section VI concludes and states some future works.

## II. Related works

### A. City Surveillance and Protection

City surveillance and protection is a hot topic with quite a lot of literature behind it. In recent years, new trends in terms of security and safety arise with the arrival of smart spaces, towards cyber security phenomenon. Nowadays, video surveillance technologies are appealed more and more. Currently, video surveillance systems are deployed in a wide range, for example, intrusion detection and traffic surveillance [3], [4].

Reference [5] declared that by monitoring people's actions, it is possible to determine a violent action, and more identify and recognize involved people. Reference [6] suggested using infrared video to track and detect pedestrians at night. Reference [7] proposed a framework using video sequences to detect people carrying objects (objects are not allowed in restricted areas). A crowd behavior classification algorithm is proposed by [8] for safety reasons.

GIS tools have been widely used for ensuring public safety. Reference [9] proposed an event-based emergency video surveillance system with a GIS platform to allow the access to real-time video data of an accident scene in a fast and efficient way. Reference [10] shows the utilization of GIS to identify the optimal number and location of observers (e.g., cameras) in order to ensure complete visual coverage of a geographical zone.

### B. Big Data issues and related technologies

One of the solutions proposed to analyze Big Data is distributed computing as suggested by [11]. Also, the Hadoop Distributed File System [12],[13] is designed to store very large data sets reliably and to stream those data sets at high bandwidth to user applications. Cisco IT built a storage cloud service called S-Cloud (Cisco IT Case Study) to store, manage, and protect globally distributed unstructured data. Cloud-based stores [14] like MongoDB [15], which is an open-source document database that provides high performance, high availability, and automatic scaling) support the storage of a huge volume of data and allow elastic scaling for handling the unexpected load. The concept of delegating jobs may ensure a good load balancing and minimize the traffic to the main central-database. However [16] proposed a good approach that investigates the problem of cloud service reliability. The approach tries to provide an enhancement solution for minimizing network and storage resources used in a cloud data center. Sure, within a global city surveillance solution, like CityPro, the above approach should be solicited at the stage when dealing with service consistency and reliability.

However, solutions like [17], which analyze video data and detect abnormal aspects and give alerts, could also be used and integrated. However, the experiment of Trento city in Italy shows that associating big-data and open-data [18] could positively impact the everyday life of citizens.

### C. Data Integration from Heterogeneous Sources

Data integration in the purest sense is about carefully and methodically blending data from different sources, making it more useful and valuable than it was before. IBM provides a strong definition, stating “Data integration is the combination of technical and business processes used to combine data from disparate sources into meaningful and valuable information.” It’s about making the data comprehensive and more usable [19].

Reference [20] states different types of traditional integration techniques including manual data integration, middleware data integration, data virtualization integration approach and data warehouse approach of integration. It also discusses the challenges that come along with data integration including design and implementation challenges.

Reference [21] discussed integrating data from different sources by XML and XQuery using a middleware known as Mediator. The approach claims to offer fast querying on heterogeneous and distributed data sources while benefiting from caching metadata for later use. Reference [22] also discussed the use of mediator architecture for integration. However, they used a different ontology which is RDF and SPARQL based. Most of the recent approaches in this domain are actually ontology-based with no recent studies in these terms.

### D. Federated Databases

A federated database is a system in which several databases appear to function as a single entity. Each component database in the system is completely self-sustained and functional. When an application queries the federated database, the system figures out which of its component databases contains the data being requested and passes the request to it [23].

In a homogeneous environment, the federated database may be composed of a heterogeneous collection of databases, in which case it lets applications look at data in a more unified way without having to duplicate it across databases or make multiple queries and manually combine the results [24].

Reference [25] identifies the requirements of a Federated DW System and proposes an architecture supporting the tightly coupled integration of heterogeneous data marts into a global, logical schema. In order to enable the processing of queries in the federation, they provide a Dimension Algebra (DA) and Fact Algebra (FA) to define the mappings between the global and local schemas.

### E. Star-Schema Database

In data warehousing and business intelligence (BI), a star schema is the simplest form of a dimensional model, in which data is organized into facts and dimensions. A fact is an event that is counted or measured, such as a sale or login. A dimension contains reference information about the fact, such as date, product, or customer. A star schema is diagramed by surrounding each fact with its associated dimensions. The resulting diagram resembles a star. Star schemas are optimized for querying large data sets and are used in data warehouses and data marts to support OLAP cubes, business intelligence and analytic applications, and ad-hoc queries [26].

Reference [27] demonstrates how a star schema can be used to model complex data warehousing. It gives multiple examples where star schema is used to model not only complex but also multidimensional data warehouses.

## *F. Data Summarization*

Summarization is a key data mining concept which involves techniques for finding a compact description of a dataset. Simple summarization methods such as tabulating the mean and standard deviations are often applied for exploratory data analysis, data visualization, and automated report generation. Clustering is another data mining technique that is often used to summarize large datasets [28].

Reference [28] suggested two approaches to data summarization which are by compressing data into informative representation. The first approach presented is by using frequent item sets while the second is by using the BUS algorithm. While this article presented those two approaches everything else in this domain was basically concentrating on summarization in data mining.

# III. THE CITYPRO COLLABORATION PLATFORM – AN OVERVIEW

CityPro is a full-collaborative platform playing the role of "dedicated decision support operational center". The platform simulates a distributed collaborative platform and applies an event-driven collaborative process (Fig. 1), it consists, on one hand, of a core system associated to a specific data-repository which is intended to emerge intelligent data, and on the other hand, consists of a set of standalone computer systems seen as computer-partners (collaborators, and or consumers).

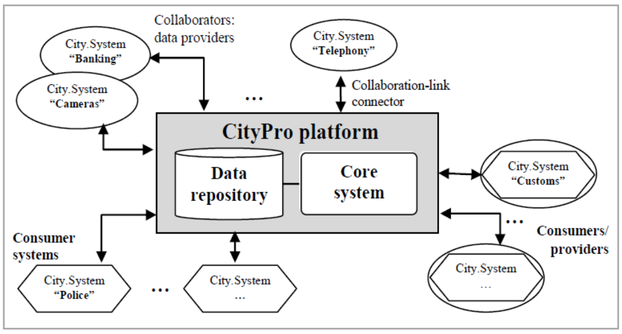



Fig. 1. CityPro's target system

1. Central system; the core system
2. Partners: Data providers (collaborator-system) - Consumer systems
3. Connectors

## *A. The Central (core) system*

The cornerstone of the system; it consists of a set of dedicated software components/engines. It applies a devoted business process and deals with complex events based on predefined patterns. The core system also supports an integrated Big Data repository.

The system is intended to deal with complex events. It emerges data-mining analysis techniques, advanced online monitoring and tracking technologies, and GIS geo-localization and visualization capabilities. It also supports and performs protected and secured collaboration tasks. The core system incorporates mainly a set of specialized components/engines consists of a big-data repository (warehouse) and **(Fig. 2).**

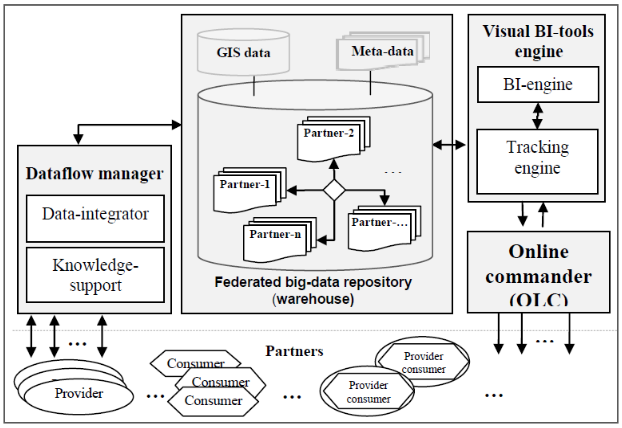



Fig. 2. CityPro's Central System (Core) Architecture

**Big-Data repository (warehouse) -** This component stores (federate) gathered information from partner systems. We opted for a federated star schema-like data model; it is an extended federated schema. Data coming from providers are summarized, integrated (adapter pattern/ETL like), etc., and is constantly planned to be enriched with the appropriate data.

**Dataflow Manager** – This engine transmits data between the core system and the registered partners. Data exchange is made through the following automated methods:

1. **Periodically:** data arrives continually from providers. The Repository should generate and emerge up-to-date (against applied periods) summarized information.
2. **On the fly (event dependent):** The central system asks providers for an instant detailed data to perform an emergent task. Also, the consumer may ask for instant data from the central system.

In both cases, strong and secure data exchange policies are applied, they set-up what data are allowed to be exposed to and what security protocols used in the data transfer process.

**Online Commander (OLC)** – It is an added tool, this component/engine acts as "City's Operating Room"; it plays a chief role in terms of surveillance intelligence. It applies an event-driven communication style and interacts with partners; and both provider and consumer external systems. The OLC symbolizes the cornerstone of the CityPro operational collaborative business process.

## *B. Partners*

**Data Providers (Collaborator-System)** – It consists of a set of domain-specific independent (standalone) systems that co-exist worldwide and within the considered territory/area. They mainly include both governmental and non-governmental systems such as police-core departments, customs, fire-stations, hospitals, banks, shop-centers, etc. These systems operate separately and continually recording granular/detailed data; each system has its own and specific business process.

**Consumer Systems** – These are decision support systems that are totally independent of each other and benefit from the 'knowledge-base' supported by the central system, e.g., police-core departments, governmental operational rooms. In return, they also play the role of 'knowledge-providers'.

## C. Connectors

Dedicated links are materializing the collaborative inter-relationships between the CityPro components. They mainly consist of ETL like dedicated data exchange automated protocols based on 'adapter pattern' [1].

# IV. SMART REPOSITORY (SMARTREP) SOLUTION

When the idea of CityPro was discussed they concentrated on the whole architectural solution by considering components and challenges, formal protocols, and scenarios, etc. However, they did mention the necessity to deeply investigate the federated big-data solution.

Our goal in this approach concentrated on transforming the Big Data collected from the partners into smart valuable data. While doing so we had to take into consideration data security and privacy, data variety, and data volume.

**Data privacy and security** – A governmental issue that would abide companies and departments to send the data by a legal contract.

**Data heterogeneity (variety)** – It is the most complicated issue in this approach. Data cannot be subjected to unification since it is produced by nonrelated sources (like banks and cellular networks). As an example, it is impossible to request data from cellular network providers using the bank ID of an individual. However, we can use that ID to get different details like a name or a civil number which will actually help in gaining access to the data present in the cellular network. That is what we are basically going to do but in a more organized and sophisticated way. Our approach is going to benefit from the metadata that it is going to create in order to access different partners and link them together. The main idea that would facilitate and reduce the heterogeneity is what we are going to call a **"Global Profile"**.

**Data volume** – We are focusing on data summarization to have this problem under control. The received data passes through three stages of summarization (Fig. 3) each having a certain identifier. These stages will reduce the amount of data extremely while maintaining its benefits and keeping it easy to access, and interpret the steps of the infamous integration method Extract, Transform, and Load (ETL).

To further explain this process we are considering as an example the presence of three different partners which are the banks, hospitals and telecommunication companies.

**Manual Feature Selection** – In the first stage of this process, we select manually (Extract) from each partner type the features (attributes) that are the most convenient for our platform's purpose. This selection is done according to a legal contract made between the partner and the system. Each partner is then granted an ID (referred to as PID) and added to the partners' metadata. The data encapsulated within its source's ID is then sent to stage 2 for the first integration. By performing this step we started the process of data summarization, by reducing the number of attributes we are going to receive from each partner, thus minimizing the size of each record.

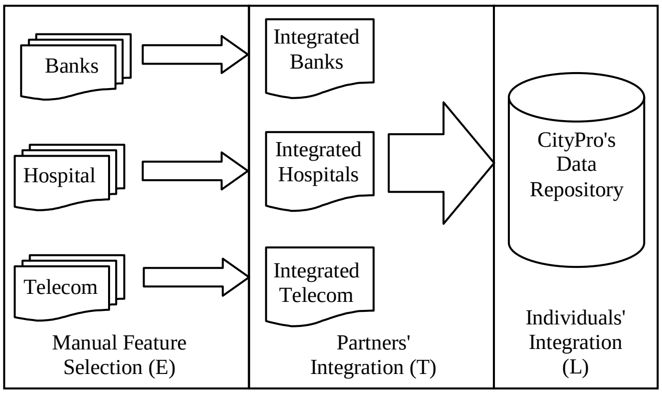
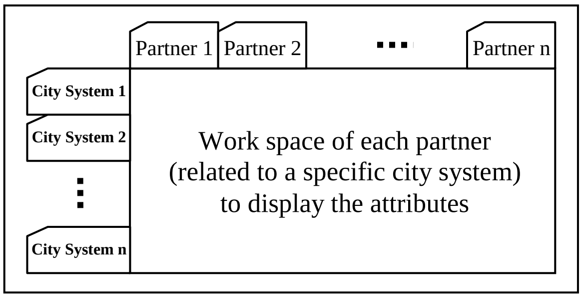



Fig. 3. Three Stages of Summarization

**Partners' Integration** – After selecting the appropriate attributes from each partner, the records received are all stored in one table. At this point, each record belongs to an individual in the city. Passing on the records one by one, each individual would get a global ID (referred to as GID) and be added to the individuals' metadata. In the created table, the primary identifier for each record would be the PID, GID, and date of the record (Transform). The goal of this stage is to improve the accessibility to the data by storing it in a smart way.

**Individuals' Integration** – The final step is integrating the data in the facts table of the repository (Load). The data delivered in stage 2 was divided according to the partner's type in separate tables. In this stage, our main identifier will be the date and GID. So to put it in simple words, a single record would be the integration of all the actions an individual has made in all partners on a specific date. The records would be pivoted to have a specific format that would demonstrate the records (TABLE I. ) in a beneficially visualized manner.

TABLE I. INTENDED TABBED VISUALIZATION OF PARTNER'S RECORDS IN SMARTREP




The treatment of these basic issues will not only help in integrating the data in the federated data-repository but will also increase the simplicity in accessing data at all times. In brief, we would be moving from Big Data to smart data.

## A. Architecture

The work is entirely based on a star-schema federated data repository component (Fig. 4).

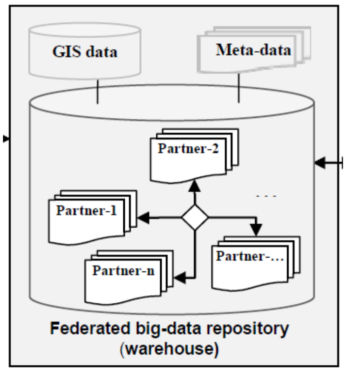



Fig. 4. Our position in CityPro

In the repository, we will have a single table (Axis) to represent each type of partner, i.e. there is no dimension table for each bank present in the city. The data produced by all the banks will be integrated into one dimension table. This is where the metadata would be used.

The global profile mentioned earlier appears here. Each partner will have a global ID set by the system and stored along with all its details in the partners' metadata table. The same idea is applied to individuals' metadata.

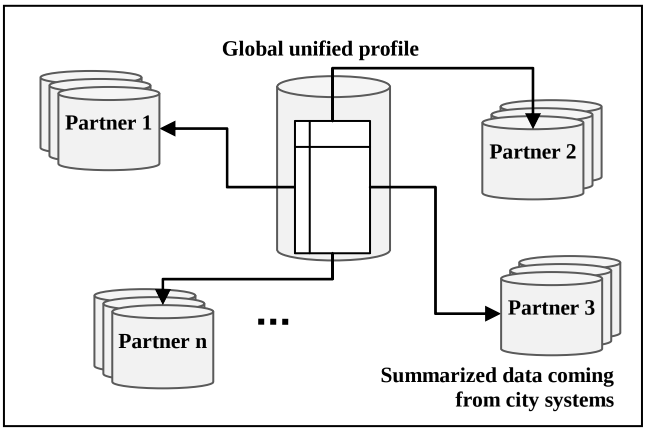



Fig. 5. Federated Star-Schema Representing Partner's Axis

Fig. 5 presents the general architecture followed in the SmartRep of the solution where the axes are related to different omnipresent systems in the city. Taking the axis showing the label partner 1 as an example all the partners on this axis (partner 1 and those that follow) are collecting data from multiple city systems that belong to the same type (banks for example). This way of handling data allows the system to demonstrate all the data from similar system types on one axis which will facilitate the visualization of data and querying when needed. I.E if we need to find the data related to an individual in multiple banks it is enough to access the axis related to banks which would deliver all the related data from all the banks present in the city.

At the same time requesting data from different partners related to one individual is also possible using the adapted architecture. The data integrated to the axes from similar partners is later on integrated to the center of this repository (which is known as a fact table in the star schema database). Acquiring data from the fact table grants access to an individual's record at a certain time in all partners present on all the axes of the schema. This data is in the considered as a summary of the individual's actions in the city at a specific timing. The adaptation of this architecture in real life is discussed in section V**.**

## V. SmartRep in Practice Typical Case Study and Scenario

To apply this architecture in real life let's consider having an individual accessing different systems in the city (Fig. 6). Each system naturally grants this individual its own unique ID. As an outsider seeing these diverse IDs without any external information will not relate them to the same individual unless this information was given to us willingly. At this point SmartRep takes action. All of the data collected from individual partners is integrated to the specified axis as shown in Fig. 6.

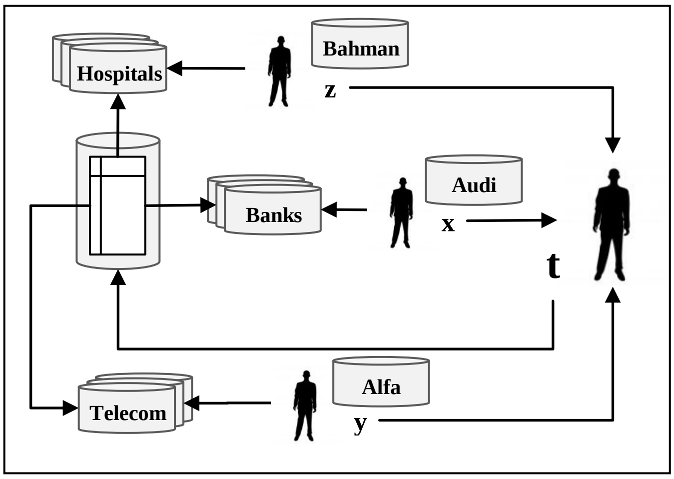



Fig. 6. SmartRep in Practice

At the same time, a global profile is created for this individual using a new global ID set by the SmartRep (**t**), that acts as an integration of all the previous IDs present (**x-y-z**). The global profile of the individual is then used to integrate all the related data and store it in the fact table (Fig. 6). This ID would be used to access all the data related to this individual in SmartRep instead of using his multiple IDs. This data is then visualized from the global unified profile.

TABLE II. Visualization of Individual **T** in Global Unified Profile

MTC | **Alfa** | +

| Hospitals | Date | GID | PID | ID | … |
| --- | --- | --- | --- | --- | --- |
| Banks | 11-11-2011 | a | 415 | d | |
| | **01-05-2014** | **t** | **123** | **y** | |
| **Telecom** | 29-10-2018 | l | 789 | r | |
| + | | | | | |

### A. *Traditional Systems vs. CityPro's SmartRep*

Integrating the specified attributes in the fact table and giving them unique identities, facilitated the tracking process enormously. To prove our claim, consider the scenario of an investigation about a suspect.

If we wanted to use the old traditional way, (Fig. 7) first we have to ask the partners if they approve giving us the details. After studying this request, there is always a possibility of rejection but let's suppose that they approved it. The partner then receives that suspect's ID sends it to its database and queries related data. However, it might just happen that this particular partner doesn't have any data related to this specific ID (take a bank's branch as an example). Logically there is nothing that abides an individual to have an account at a certain bank rather than the other. Therefore, the result of the search could turn out both positively and negatively; in both cases that database would reply to the partner which will, in turn, respond to the initial requester. Keep in mind that after this long process we might have or might have not gathered data from a single partner depending on their responses.

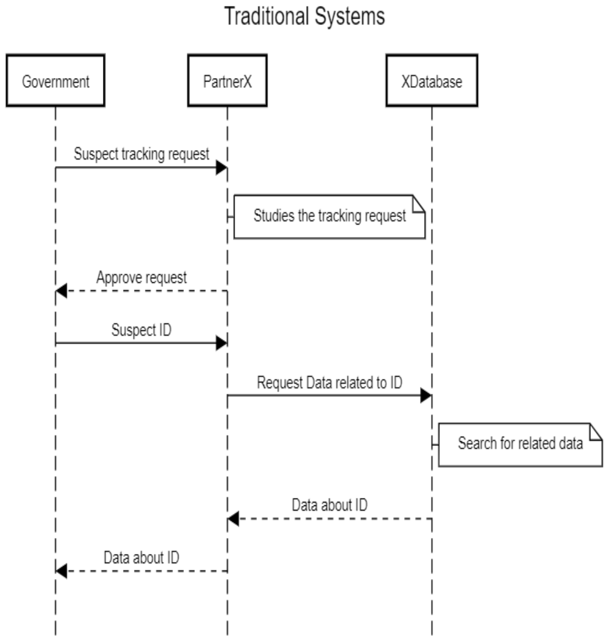



Fig. 7. Traditional Sequence Diagram

On the other hand, using CityPro's SmartRep (Fig. 8) to gather data about a suspect required a simple request by sending an individual's identifier. The identifier could be his name, phone number, civil number, etc. since all these details are found in the metadata of the repository. CityPro simply links this ID to its suitable GID and then starts to gather all the integrated data found in the repository that is related to this GID. Note that by these simple steps, the data was gathered from all the partners in the city.

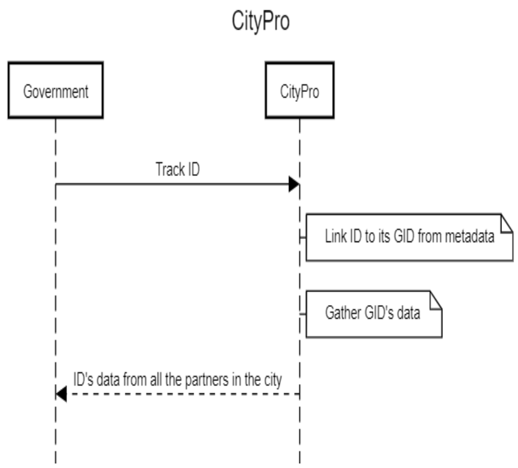



Fig. 8. SmartRep's Sequence Diagram

## VI. Conclusion, Observation and Future Works

Working on the presented domain is only the beginning; it needs more development and treatment of different aspects to become effective. The challenges presented by [1] model a path towards further studies within the topic. Addressing them is a must in order for this project to evolve and come to life. The adaptation of this platform in any modern city might be the key to solving various problems and obtaining a safe life. Whenever it is executed it shall be a great advantage for the national security of any country and lead to its development enormously.

Our position regarding all the previous work done is very advanced. For one, we integrated the concept of a federated database into a star-schema model which is, based on our research, a new concept that one had yet executed.

CityPro, as stated in [1], is prior to most traditional surveillance systems. It can be seen as a system that has combined all the known surveillance system into its platform.

When it comes to data integration from heterogeneous sources, most approaches followed are ontology-based. They usually require knowing the structure of the databases they are dealing with on the other side or abide by the data producers to follow a specific format. However, as CityPro we promised the least possible interference with the data providers and adopted a method that deals with all data type with the least possible alterations. Regarding data summarization, we used ETL in the process of summarizing data, which is also considered as a new concept in this domain. Achieving our foremost goal which is to transform, by benefiting from all those techniques, the Big Data produced by all the partners into smart summarized data.

We hope to implement CityPro as a real platform in the near future where we would have access to streams of real data in order to test our approaches full capabilities. Later on, we plan to increase the intelligence of the system as a whole and the repository in precise. Possible feature selection algorithms might interfere in the manual feature selection to make the process fully automated. This automation would be extremely helpful when we start to incorporate machine learning techniques in the data analysis to guarantee more accurate results. Moreover, we intend to continue working on this approach in order to seek any possible optimization. In addition, we will also start addressing the business processing issue, and integrate the two solutions together.

Finally, what we opt for is to extract CityPro's hidden powers in order to benefit from it in an optimal way. That's why we shall try to address all the issues that might face its happening. In the end, our ultimate goal is to create a utopian collaborated surveillance system called CityPro.

## Acknowledgments

This work is supported and funded by the Department of research at the Lebanese University, Lebanon.

## References

[1] Mohamed Dbouk, Hamid Mcheick, Ihab Sbeity. (2017). CityPro: city-surveillance collaborative platform. Int. J. Big Data Intelligence, IV (3), 161-170.

[2] Bawany, N.Z., and Shamsi, J.A. (2015). Smart city architecture: vision and challenges. International Journal of Advanced Computer Science and Applications (IJACSA), VI (11), 246-255.

[3] Kalirajan, K. and Sudha, M. (2015). Moving object detection for video surveillance. The Scientific World Journal, an Open Access Article under the CCAL, MMXV.

[4] Bhajibhakare, M.M. and Deshmukh, P.K. (2013). To detect and track moving object for surveillance system. International Journal of Innovative Research in Computer and Communication Engineering, I (4), 945-949.

[5] Hancke, G.P., Silva, B.C. and Hancke Jr., G.P. (2013). The role of advanced sensing in smart cities. Sensors, XIII (1), 393-425.

[6] Wang, B., Ye, M., Li, X., Zhao, F. and Ding, J. (2012). Abnormal crowd behavior detection using high-frequency and spatial-temporal features. Machine Vision and Applications, XXIII (3), 501-511.

[7] Damen, D. and Hogg, D. (2012). Detecting carried objects from sequence of walking pedestrians. IEEE Trans. Pattern Anal. Mach. Intell., XXXIV (6), 1056-1067.

[8] Wang, J., Chen, D., Chen, H. and Yang, J. (2012). On pedestrian detection and tracking in infrared videos. Pattern Recognition Letters, XXXIII (6), 775-785.

[9] Xiong, X., Wang, B. and Wang, D. (2009). Research of event-based emergency video surveillance system. International Workshop on Information Security and Application (IWISA 2009). Finland.

[10] Rana, S. (2005). Use of GIS for planning visual surveillance installations. Procs. ESRI Homeland Security GIS Summit. Denver, Colorado.

[11] Jacobs, A. (2009). The pathologies of Big Data. In Communications of the ACM, 36-44.

[12] Hadoop-DFS. (2016). Hadoop Distributed File System. Retrieved from http://hadoop.apache.org/docs/current/hadoopproject-dist/hadoop-hdfs/HdfsDesign.html

[13] Shvachko, K. K. (2010). The Hadoop distributed file system. Proceedings of the IEEE 26th Symposium on Mass Storage Systems and Technologies, 1-10.

[14] Dey, S., Chakraborty, A., Naskar, S. and Misra, P. (2012). Smart city surveillance: leveraging benefits of cloud data stores. IEEE 37th Conference on Local Computer Networks workshops, 868-876.

[15] MongoDB. (2016). Retrieved from http://www.mongodb.org/manual

[16] Zhu, Y. and Zuo, J. (2015). Research on security construction of smart city. International Journal of Smart Home (IJSH), 197-204.

[17] Calavia, L., Baladrón, C., Aguiar, JM., Carro, B. and Esguevillas, A. (2012). A semantic autonomous video surveillance system for dense camera networks in smart cities. Sensors, 10407-10429.

[18] Molinari, A., Maltese, V., Vaccari, L., Almi, A. and Bassi, E. (2014). Big Data and open data for a smart city. White Papers from the IEEE Smart Cities Inaugural Workshop, Trento, Italy.

[19] GlobalScape. (2017, December 6). 5 Types of Data Integration You Need to Know. Retrieved from GlobalScape: https://www.globalscape.com/blog/5-types-data-integration

[20] Kumar Gauraw. (2015, February 27). Data Integration Ninjas. Retrieved September 22, 2018, from http://www.dataintegration.ninja/data-integration-techniques-and-its-challenges/

[21] Gardarin, G., Mensch, A., Dang-Ngoc, T.-T., & Smit, L. (2002). Integrating Heterogeneous Data Sources with XML and XQuery. International Workshop on Database and Expert Systems Applications. Dexa.

[22] Jinpeng Wang, Jianjiang Lu, Yafei Zhang, Zhuang Miao, and Bo Zhou. (2009). Integrating Heterogeneous Data Source Using Ontology. JOURNAL OF SOFTWARE, IV (8), 843-850.

[23] Rouse, M. (2011, February). Virtual (federated) database. Retrieved from Search Oracle: https://searchoracle.techtarget.com/definition/virtual-federated-database

[24] Shavit, Y. (2008, July). TechTarget. Retrieved September 2018, from https://searchitchannel.techtarget.com/feature/What-are-federated-databases

[25] Berger, Stefan, and Schrefl, Michael. (2008). From Federated Databases to a Federated Data Warehouse System. Hawaii International Conference on System Sciences.

[26] Rouse, M. (2010, September). Star Schema. Retrieved from Search Data Management: https://searchdatamanagement.techtarget.com/definition/star-schema

[27] O. Boussaid, Adrian Tanasescu, Fadila Bentayeb, and Jérôme Darmont. (2007). Integration and dimensional modeling approaches for complex data warehousing. J Glob Optim (37), 571-591.

[28] Varun Chandola and Vipin Kumar. (n.d.). Summarization - Compressing Data into an Informative Representation.