Advanced Data Analytics and Visualisation for the Management of Human Perception of Safety and Security in Urban Spaces

Panos Melas, Gianluca Correndo, Lee Middleton, Zoheir A. Sabeur

Advanced Data Analytics and Visualisation for the Management of Human Perception of Safety and Security in Urban Spaces

Panos Melas, Gianluca Correndo, Lee Middleton and Zoheir Sabeur

{pm,gc,ljm,zas}@it-innovation.soton.ac.uk
University of Southampton IT Innovation Centre, Electronics and Computer Science, Faculty of Physical Sciences and Engineering, Southampton, United Kingdom

Abstract. The genesis of this work began during the DESURBS[1] project. The scope of the project was to help build a collaborative decision-support system portal where spatial planning professionals could learn about designing much more secure and safer spaces in urban areas. The portal achieved this via integrating a number of tools under a common, simple to use, interface. However, the deficiencies in the project became apparent with subsequent development. Many of the open data employed changed format while applications were increasingly custom built for a single dataset. In order to overcome this a system called KnowDS was redesigned. The essence of the new design includes decoupling acquisition, analysis and overall presentation of data components. The acquisition component was designed to snap-shot the "data providing methods" and query data provenance in a similar way to a source code repository. The analysis component is built under a number of modular tools with a common interface which allows analysis to build in a plug&play approach. Finally, the data presentation component is where the custom logic goes. Under such design approach, the building of future applications becomes less challenging. As a consequence, two case studies using the new framework were considered. Firstly, a UK crime web-browser which allows data analytics performances at various granularities of crime types while correlating crimes across various UK cities has been achieved. Secondly, a mobile application which enables to generate reports on citizens' perception of safety in urban zones has also been developed. The two applications were efficiently built under the new design framework; and they clearly demonstrate the capacity of the new system while they actively generate new knowledge about safety in urban spaces.

Keywords: safety perception, urban security, data analytics, visualisation, open data

1 Introduction

Large volumes of open observation and contextual data concerning urban spaces have become more accessible for processing in recent years. The open data can be potentially aggregated for better usage and improvement of existing decision-support sys-

[1] DESURBS: Designing Safer Urban Spaces: http://www.desurbs.eu

tems, specializing in the dissemination of safety related issues in cities to the public [1]. The potentially useful data aggregation exercise starts with the use of the various open data sources which are provided as part of the open government movements globally [2]. Additionally, social network data, or so-called social sensing [3] data can be used [4, 5]. Furthermore, survey information from custom written mobile applications can be employed. The aggregation with private (closed) data such as local authority data or data from telecom providers can be employed [6]. The heterogeneous nature of the data can be improved via a data normalisation process. This converts all data to a common schema. However, the data is semi-structured and therefore concepts from NoSQL databases can be adopted [7]. Thus, all sources of data are analogous to a function which maps from space-time to a series of values. Data collection is performed using a periodic process which polls the data source for any changes. Because the data sources are potentially changing, it is important that the system not only stores the latest snapshot but also changes in data. For both operations, i.e. data collection and data normalisation, different filters/adapters are developed in order to fetch and normalise data. For example, the anonymisation of sensitive data might take place during normalisation. The system should be able to handle multiple adapters in order to fetch and normalise data accordingly.

An advanced Knowledge Discovery System (KnowDS) that can aggregate open data provided on the cloud has been developed in this study following the progress made in the DESURBS project [8]. The system partially structures the data in terms of a common schema and provides methodologies for updating and analysing urban data. Furthermore, the use of common external API allows applications to be built efficiently on top of the system. This paper is structured as follows. Firstly, a technical overview of KnowDS is presented with highlights on the encountered significant components and challenges. Secondly, a set of concrete examples of the framework in operation are described. Finally, conclusions are drawn along with future work and recommendations.

2 KnowDS Design Overview

It is becoming increasingly common for data platforms to bring together open datasets and perform visualisations upon them. Such systems have three main components. Firstly, there is the acquisition system which downloads the data from an external API and stores in a custom database. Secondly, there is the processing system which takes the data form the database and converts it ready for display. Finally, there is a client application, typically web-based, which embeds the data and displays it. As they are designed for a specific purpose and each steps feeds to the next, it is difficult to repurpose the data easily and efficiently. As an example a change in the external API may break the system completely. This issue was experienced in the early ongoing work phase of the DESURBS project.

Building upon such experience in DESURBS, the KnowDS framework was developed. The framework is divided into three components: Data collection and normalisation; Data processing and analysis; and applications. These are illustrated in Fig. 1

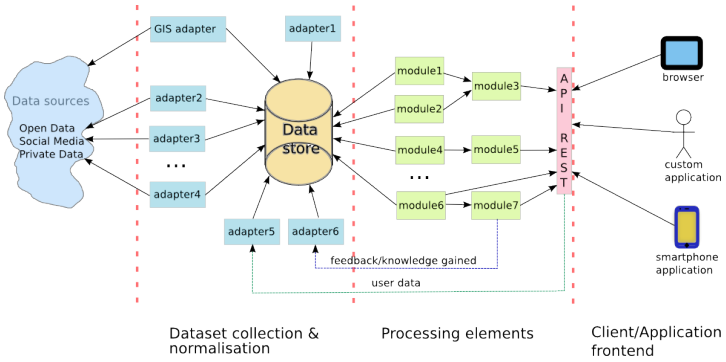


Fig. 1. Overall Design of KnowDS

above and delineated by the dashed lines. Each component is composed from individual modules, i.e. adapters and processing elements. Processing elements can provide a standardised API that frontend client applications and employ to interact with the system. Both adaptors and processing elements have clear interfaces while they perform a single specified task. Nevertheless, more complex tasks and analyses can be achieved by combining a set of processing elements. The overall system design is open and flexible to extend and accommodate new data sources as well as new data processing elements for performing knowledge discovery. The core implementation is primarily based on open source technologies and languages. The core technologies include Python, PostgreSQL, PostGIS, Django and Apache. The various components are discussed with more details in the next sub-sections.

2.1 Data Collection and Normalisation

The purpose of the *Data Collection and Normalisation* component is to gather and extract data of interest from various data sources. This has a similar purpose to conventional ETL (Extract-Transform-Load) [9, 10] systems which are common in data warehouses. In the extract process consumes web APIs (but more generally any data source). The transform serves to add structure to the original data which is often unstructured and heterogeneous. In contrast to ETL systems the reconstruction of the history of the data as well as current snapshot is enabled. For example, it is important to know that a specific API splits one field into two at specific date and provide snapshots of the data before and after the change. In order to achieve this, a method of tracking the changes to the data, similar to those employed by version control systems, has been purposely implemented. The KnowDS data store is composed of a combination of databases, directory structure and a storage archiving system. The current implementation of the data store employs a combination of databases and an auxiliary hierarchical directory structure to store binaries and data blobs. In order to realise the above vision, the nature of the datasets available (i.e. from data.gov.uk) to classify their nature was studied. In all cases the dataset was modelled as a functional

Dataset	Spatial	Temporal	ID
Twitter	×[1]/✓	✓	✓
Postcodes	✓		✓
UK Crime	✓	✓	
Weather	✓	✓	✓
Pollution	✓	✓	✓
Conservation Areas	✓		✓
Building information	✓		✓
House prices	✓	✓	
StreetView data	✓		

Table 1. Exemplar Datasets with the Common Attributes.

mapping from a spatio-temporal attribute with a resource identifier to a multi-dimensional vector.

This vision is fully documented for multiple datasets, as illustrated in Table 1 above. The fields often require conversions in order to put them under a common representation. For instance, geographical data for twitter are in WGS84 format, while the UK crime data are in Easting/Northing. Thus, the structuring process of data takes the arbitrary dataset and converts it into table format with several mandatory fields and a blob which contains the remainder of the data. It is for this reason that semi-structured data formats need to be described. The specific implementation of the component is made up of a number of autonomous modules (adapters) that are capable of accessing the source data, conversion, structuring, and storage. Specific adapters are lightweight and reusable since they are coded in a common pattern. Typical activities with the adapters include fetching the data from the source (one shot or periodic), data normalisation and anonymisation (of sensitive data). An example of data conversion processing can be illustrated through a relatively simple geographical dataset. Urban spaces can be simply defined as geospatial polygons. However, their representations differ among local authorities. Among the representations commonly used in the UK include:

- **Cities**: A boundary dataset containing polygons for each one of our cities. Defined at the level of government.
- **Neighbourhoods**: This is defined by the census Middle Layer Super Output Areas (MSOAs) are population-wise similar (~7400). Provided by Ordinance Survey.
- **Postcodes**: The highest level of resolution. Often down to the granularity of a few households. This is provided by the postal service.

These each are at a different zoom level and size. While some of the boundaries match (city and neighbourhood) together, others do not (postcodes). This means that a typical postcode can be contained in 1 or more neighbourhoods. Additionally, different schemes use different coordinate systems. WGS84[2] was used as a common coordinate scheme for our geographical coordinates. This means that on ingest the conver-

[2] http://earth-info.nga.mil/GandG/wgs84/index.html

sion to the WGS84 reference coordinates will be required. In order to cope with changes in data structure, many sorts of transformations that happen over the lifetime of datasets were addressed. The mapping relationship which was described earlier can be considered as data model. The model transformations are thus: *Creation, Modification and Deletion*. In order to handle such transformations, the history of the dataset which can be used for replay has been maintained. Hence, the restoration of previous states of the data even when the dataset has changed can be achieved. An example of the need for changes can be seen by studying the history of a particular API. For example the UK Police crime data API. In 2013, there were several breaking changes to it. Firstly, a refresh of the crime data meant that several crime types were reclassified. Secondly, the ability to obtain crimes in a specific Police neighbourhood (in Easting/Northing) was removed and replaced by street-level information (WGS84). As a result, several police administrative regions were merged. The reclassification of the crime types required the addition and removal of fields to the data model. The API change can be dealt with via the addition of a new model for the data. The merging of regions can be solved via modification of the data in the model.

2.2 Dataset Processing Modules

Whilst the purpose of the previous module was to collect and structure the data, the data processing module performs operations on the data for subsequent use by an application. The processing takes place in the central portion of **Fig. 1**. There are a number of specialised modules which perform processing of data which are contained within the dataset and provide output. Like the data collection adaptors, the modules are written under a common pattern. They specifically manage the access to the dataset and analyse it for a given specific task. Once analysed the output can be exposed via an API or passed to another module for subsequent processing. In this sense the modules can be classified in terms of whether the goal is to analyse data or aid in client applications.

The analysis takes place depending on the current client requirements and needs. In this model the back-end acts as a broker for data which is passed to an analytics engine. The analytics engine also stores data internally in a database. However, the data is in processed form from the broker. There are a number of different data processing

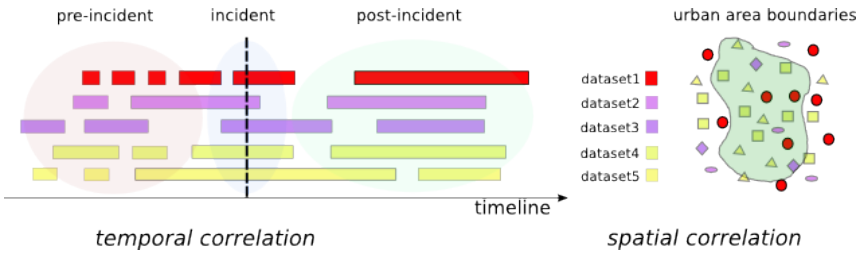


Fig. 2. Timeline, or location based analysis combining a heterogeneity of data sources revealing a pattern

techniques that can be applied by the analytics engine. They are:

- *Simple analysis of data* (From a particular resource).
- *Combining data from different resources* (e.g. Timeline analysis on data from different sources, identification of patterns, behaviour.)
- *Incident timeline analysis* (Which includes combined data from available resources at a given time of an event.)

In particular, the analysis of data can split into pre-incident, during-incident, and post-incident analyses (see **Fig. 2**). Spatial analysis can also be applied on predefined urban spaces using multiple datasets as shown in **Fig. 2**. Higher level analysis, or customised user defined analysis should be facilitated, i.e. in terms of describing an analysis as a workflow/process that can be loaded and run by the system. The analysis and study of identified patterns/behaviour, which are external to the system, might require input from knowledge domain experts.

2.3 The Application Front-End

At the frontend level, the business logic for the specific application should be contained. This is specific to the application and/or visualisation and does not enable the degree of code reuse as exhibited for the previous components. No analysis of data takes place at this level, however, data processing for visualisation or conversion to another format might be required. The connection between the clients and platform is via an established API such as HTTP/REST. Many applications will require some processing of the raw data, while this is triggered via the API. The results are cached within the data-store for subsequent use. This is illustrated in the right-most component of **Fig. 1**. Some examples of the sorts of modules that can be provided are: *Mash-up visualisations* (where data from different datasets are normalised on similar coordinates); *Custom data exporting*; and *Entering of user contributed data*.

3 Case Studies

Two case studies using the KnowDS framework are presented in this section. The first case study is an application which extracts data from the open UK crime API and visualises it using spatial analytics with Ordinance Survey data. The second case study is a mobile application that enables users report on their perception of safety at urban spaces which they frequent. The application shows how user contributed data can be used along with other sources of public data to generate and discover new knowledge.

3.1 UK Crime Application

This is an application that was developed for the system, while it takes into account existing work on crime analysis in the literature. The main dataset of this application

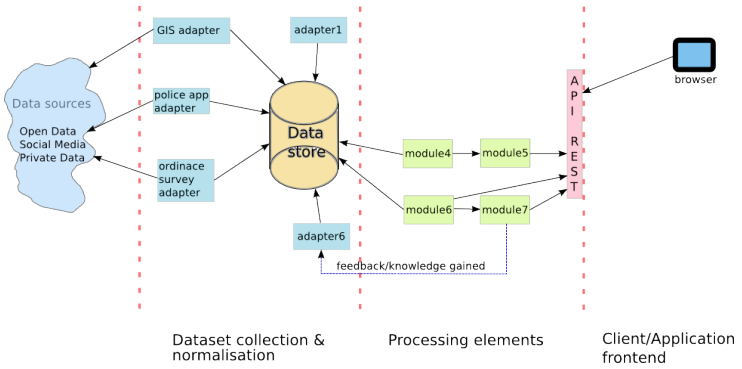


Fig. 3. UK Crime Application Design

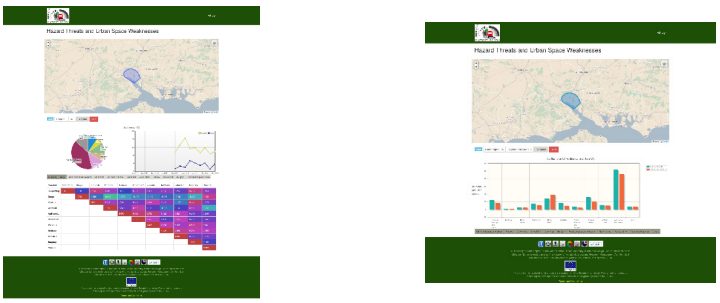
is the UK Police Crime data which is available under a public API[3]. The Police crime application examines crime data in 13 major UK cities based in the South of England, the Midlands, the North of England and Greater London. The processing tools of this application allow comparison and visualisation of different crime types at various levels of spatial granularities and zoom levels. Additionally, analytics are performed on the data to allow quantified on-demand statistical comparisons of specific urban crimes and type of crimes across UK cities. An overview of the system as implemented in the KnowDS framework is illustrated in **Fig. 3**. The main function of the adaptors in this case is to fetch data from the UK police API and store it in the database. Additionally, as the crime location data is stored in Easting/Northing conversion to WGS84 is required. Finally, the adaptor periodically downloads and updates the data (on a monthly cycle) accordingly. In order to process the data and compare at differing zoom levels, access to Ordinance Survey definitions is also required. These are ingested into the data-store via an additional adaptor.

[3] http://data.police.uk/

There are a number of distinct modules present in the system. Firstly, there is a tool that aggregates the normalised crime data for a specific crime-type and produces a "heat map" as a result. The output is archived within the data-store to reduce processing load. Secondly, there are raw crimes data which are exposed as a JSON object. This can be used to create maps on the client side. Thirdly, analytics are performed to cross-correlate various types of crime with each other. The results of the analysis are equally exposed via JSON. Finally, a simple aggregate of crimes at a given specific location (and level of zoom) is presented via a JSON object. The Police application front-end visualisation UI is a browser driven tool that can visualise various crime points and intelligent statistics on maps with cross-comparisons between different crime areas, e.g. city, neighbourhood, or postal code level. This is shown in **Fig. 4**. The data is generated dynamically from the framework and represents new investigative knowledge about urban and cross-urban crimes.

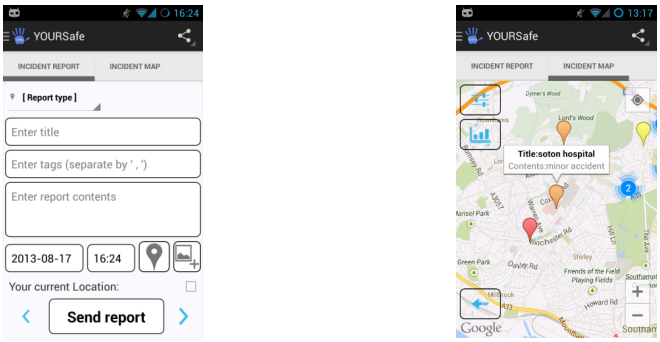
3.2 Mobile Application (*YourSafe*)

This is a native android application for reporting user safety perceptions. End users are the dataset suppliers and also consumers at the same time. YourSafe is built to allow users to report their perception of safety within urban spaces. The aim is to show that end users can become a sustainable source of information about geospatial perception of safety in urban spaces. At the same time such application can implicitly lead to making local urban spaces safer. Data protection, privacy and security is considered highly in this application under the European Data protection regulations. Device ID's are hashed and personal information is NOT stored in the system. An overall schematic of the design is shown in Fig.5. As far as the data collection is concerned, there are a number of distinct modules which are implemented. Firstly, there is an adaptor which pulls in geographical locations from Ordinance Survey. Specifically, this is re-used from the police application. Additionally, UK crime data is re-



(a) **(b)**

Fig. 4. (a) Illustration of the application showing crime data for a specific city (b) Comparison of two cities



(a) **(b)**

Fig. 6. Client application for YourSafe (a) report creation (b) visualisation

used from the application as well. The final adaptor was built in the system in order to save reports into the data-store. Optionally, a binary object such as an image, can be sent as an attachment with the report. This adaptor converts the incoming report from the API to a space-time location and stores the report alongside the user. There is very little analysis that is performed for usage in the tool. The tool provides a simple visualisation of reports and crime data at a given location. Thus a pair of modules are written to perform a spatial query to return results via JSON. In addition there are tools related to user management. Specifically, creation and deletion of users lead to updates to the underlying data store. These are called via the rest API from an external client. The Android client is pictured in **Fig. 6**. The user can create a report either online or later on via the interface in **Fig. 6(a)**. As well as the title and description, the user can add free tags to describe what is happening in terms of safety around their spaces and upload incident images. The visualisation interface (**Fig. 6(b)**) shows reports in the local neighbourhood but also Police crime reports (shown in blue). These are aggregated to an icon when there are several reports in close geographical proximities. Early participatory experimentations of Citizens has shown that the submitted

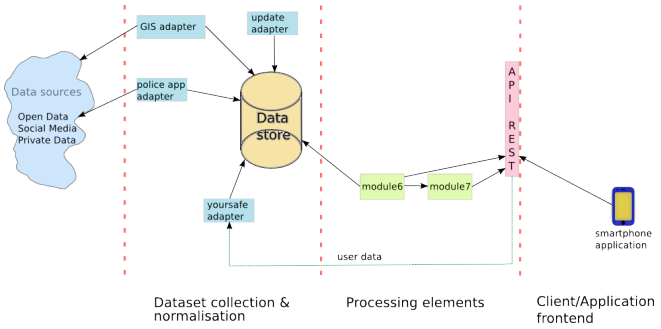


Fig. 5. The design of the YourSafe mobile application

reports are a valuable source of new knowledge. Although the number of participants in the early trial was limited, around 60% of the reports concerned safety perception narratives, 25% on alerted incidents; and 15% were on joint reporting of incidents and perception of safety in urban zones.

4 Conclusions

This paper provides an overview of the KnowDS framework for City crime data analytics. It illustrated its usefulness on two distinct applications: a UK crime web-application, and YourSafe mobile application. The essence of the design was to decouple the data gathering, data analysis, and data visualisation parts. The data gathering phase consists of a number of independent adapters which can ingest data into a data-store. Additionally, the history of the dataset is maintained and enabled to querying over historical data. This component of the framework is a data repository similar to a source code repository. The data analysis also consists of a number of modules which perform analyses on data. Processing is achieved by combining one or more of these modules. The last component is the specific application which varies depending on the application. The specific applications are very different in nature but share some common data (Police crime data). Using KnowDS allowed the mobile application to be built simply by concentrating on the differences between the two applications. Additionally, the framework demonstrated that knowledge can be mined from data content either via novel use of disparate datasets (UK Crime application); or user contributed data along existing data (YourSafe). Further development of the framework will continue in the future. More adaptors which allow more datasets to be explored will be put in place. Additionally, the range of data analytics tools will be deployed.

Acknowledgements

This work is co-funded by the European Commission under the 7th Framework programme under the DESURBS project: Grant Number 261652.

References

1. I. Vilajosana, J. Llosa, B. Martinez, M. Domingo-Prieto, A. Angles, X. Vilajosana, “Bootstrapping smart cities through a self-sustainable model based on big data flows”, IEEE Communications Magazine, Volume 51:6, 2013.
2. S. Chun, S. Shulman, R. Sandoval, E. Hovy, “Government 2.0: Making connections between citizens, data, and government”, Informaiton Policy, 2010.
3. C. Aggarwal, T. Abdelzaher, “Social Sensing”, Managing and Mining Sensor Data, pp. 237-297, 2013.

4. P. Prasetyo, M. Gaio, E.-P. Lim, C. Scollon, “Social Sensing for Urban Crisis Management: The case of the Singapore Haze”, Lecture Notes in Computer Science, Volume 8238, 2013.
5. J. Lehmann, B. Goncalves, J. J. Cattuto, “Dynamical classes of collective attention in Twitter”, WWW, 2012.
6. D. Ceolin, L. Moreau, K. O'Hara, W. Fokkink, W.R. Van Hage, V. Maccatrozzo, A. Sackley, G. Schreiber, Guus, N. Shadbolt, “Two procedures for analyzing the reliability of open government data”. International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU), 2014.
7. N. Leavett, “Will NoSQL databases live up to their promise?”, IEEE Computer, Volume 43:2, 2010.
8. A. Bonastos, L. Middleton, P. Melas, Z. Sabeur, “Crime Open data aggregation and management for the design of safer spaces in urban environments”, Environmental Software Systems: Fostering Information Sharing. ISESS, 2013.
9. H. Agrawal, G. Chafle, S. Goyal, S. Mittal, S. Mukherjea, “An Enhanced Extract-Transform-Load System for Migrating Data in Telecom Billing”, IEEE International Conference on Data Engineering (ICDE), 2008.
10. J. Henrard, J.-M. Hick, P. Thiran, "Strategies for data reengineering," Working Conference on Reverse Engineering (WCRE), 2002.