Urban Informatics: Using Big Data for City Scale Analytics

Bonsang Koo[1] Byungjin Shin[2]

Abstract: Urban Informatics, the application of data science methodologies to the urban development and planning domain, has been increasingly adopted to improve the management and efficiency of cities. This paper introduces state of the art use cases in major cities including New York, London, Seoul and Amsterdam. It also introduces recent advances in using Big Data by multi-lateral institutions for poverty reduction, and startups utilizing open data initiatives to create new value and insights. Preliminary research performed on using Seoul's open data such as building permit data and health code violations are also introduced to demonstrate opportunities in this relatively new but promising area of research.

Keywords: Urban Informatics, Big Data, Open Data Initiatives, Seoul Public Building Data

I. Introduction

Today, 54 per cent of the world's population lives in urban areas, a proportion that is expected to increase to 66 per cent by 2050. Projections show that urbanization combined with the overall growth of the world's population could add another 2.5 billion people to urban populations by 2050 [1].

This massive growth in urban density and scale will compound ongoing city challenges related to climate change, energy, infrastructure, public health, and more [2].

One of the approaches to tackle this issue is the use of Big Data technologies. Data about cities are being collected at an unprecedented scale through sensors nodes, mobile phones, smart meters, and social media. In addition, leading cities including New York, Chicago and Seoul are opening their data to the public.

With all this data becoming available, city governments around the world are starting to use data to help plan and manage their jurisdictions more effectively; i.e., to become 'smart cities.'[3] The opening of data promises new opportunities in research and practical applications including building asset valuation, city utility O&M planning, hazardous building detection, and data-driven urban planning.

Such initiatives are also not limited to the developed world, as multi-lateral institutions such as the World Bank and the UN have also caught on the opportunity and the value of using data.

This research first investigates the major initiatives in using data science technologies for sustainable urban development, referred to as 'urban informatics.' Subsequently, I introduce preliminary case examples performed on using Seoul's open data, including the use of building permit data, and restaurant hygiene violation data.

II. Research Background

A. Open Data Initiatives in Global Cities

1) New York's Geek Squad

New York, under the guidance of Mayor Bloomberg and its newly appointed 'director of analytics' Mike Flowers, set out to use its troughs of data to better manage the city of 800 million people. A problem they initially addressed was "illegal conversions", the practice of cutting up a dwelling into many smaller units so that it can house as many as 10 times the number of people it was designed for. They are major fire hazards, as well as cauldrons of crime, drugs, disease, and pest infestation.

The city needed to cover 900,000 buildings with only 200 inspectors. The city also received over 25,000 calls annually. The team used data from 19 different agencies, which included information about the age and retrofit of buildings, and also the financial status of its owners, rodent history, crime rates and more. Based on the analytics, they were able to improve the detection rate from its initial 13% per visit to 70%. This was a significant improvement and also ensured the safety of a buildings inhabitants and local fire force.

2) London Heat Map

The London Heat Map [4] is an interactive tool that allows users to identify opportunities for decentralised energy projects in London (Figure I). It builds on the 2005 London Community Heating Development Study. All information has been updated and the map is now in a user friendly format using an interactive GIS system.

The Heat Map provides spatial intelligence on factors relevant to the identification and development of Decentralised Energy (DE) opportunities, such as: major energy consumers, fuel consumption and CO2 emissions, energy supply plants, community heating networks, and heat density.

It is publicly accessible to anyone with an interest in DE. Local authorities can use the map as the starting point to

[1] Associate Professor, Dept. of Civil Engineering, Seoul National University of Science and Technology, Seoul, Republic of Korea, bonsang@seoultech.ac.kr (*Corresponding Author)
[2] Graduate Research Assistant, Dept. of Civil Engineering, Seoul National University of Science and Technology, Seoul, Republic of Korea, byungjin0826@gmail.com

developing detailed Energy Master Plans to inform DE policies in their LDFs and climate change strategies. Developers can use the map to help them meet London Plan DE policies (connection into an existing network or extending their own communal heating networks beyond their site boundaries).

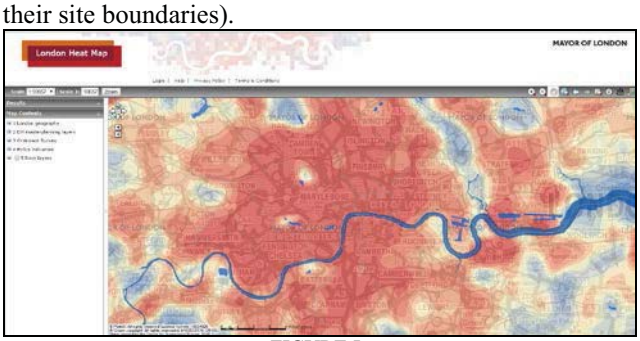


FIGURE I
LONDON HEAT MAP

3) Seoul's Owl Bus

Officials in Seoul had long searched for a transport system for low-income workers who commute late at night. Low-income workers do not make enough money to take a taxi regularly, and taxi fares are considerably higher at night. Furthermore, since low-income workers tend to live on the outskirts of the city, taxi drivers often are reluctant to go there mainly for distance and security reasons[5].

Part of the solution was the analysis and utilization of Big Data to come up with a suitable mode of transport that would serve the specific needs of late-night workers. The result was the creation of the "owl bus," which operates late into the night until five o'clock in the morning.

By utilizing only Big Data, without field surveys, nine late-night bus routes were designed, and they are currently operating between midnight and five o'clock in the morning.

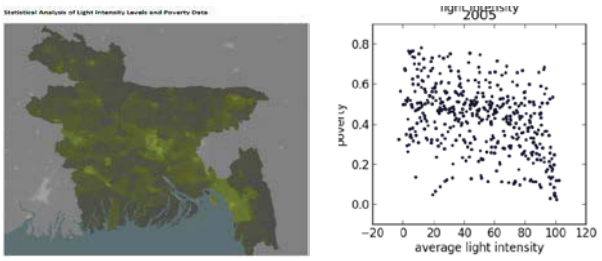
The response of late night travelers to the "owl bus" has been very positive, with 629,752 passengers who have reportedly used the new "owl bus" services.

B. Multi-lateral Institutions

1) World Bank Big Data Dive

Multi-lateral banks and the United Nations are also undertaking initiatives in using Big Data to improve global development and tackle urban issues. The World Bank recently undertook a 'Big Data Dive'[6], in which they gathered renowned data scientists and companies to explore the use of new and alternate data sources to improvement their performance and responses to global development issues.

One such research was the use of night illumination recorded from satellite imagery to identify poverty levels in a country's rural areas. Detecting and collecting poverty levels and their locations is an expensive and slow business. Using freely available satellite imagery data helped in complementing existing records and make identification faster. Figure II(a) shows night time imagery data overlaid with poverty levels for Bangladesh in 2005. Figure II(b) shows a scatter plot between light intensity and poverty levels, showing a strong correlation, which is reinforced by low RMSE values.



(a) (b)

FIGURE II
(A) NIGHT ILLUMINATION SATELLITE IMAGES OF BANGLADESH
(B) CORRELATION BETWEEN LIGHT INTENSITY AND POVERTY LEVELS

2) UN Global Pulse

The UN created a new division called Global Pulse was created with the vision to accelerate discovery, development and scaled adoption of big data innovation for sustainable development and humanitarian action[7].

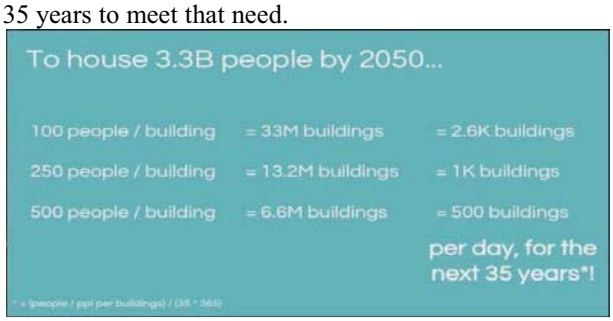
One of its cases uses social media (Tweets) to predict the price of food, oils so that it can ascertain related problems more quickly.

C. Startups

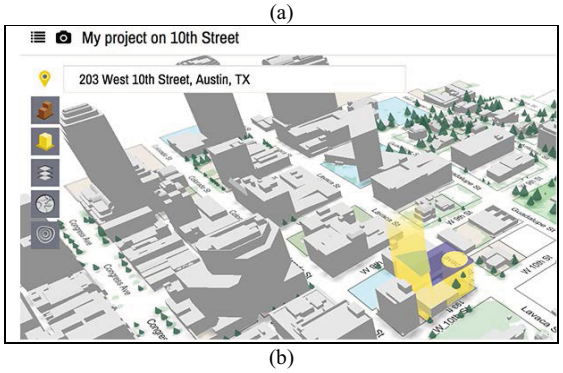
1) Google Flux

Finally, startup companies at Google have recently joined in this opportunity. Most notable is Google Flux, a platform that lets the building industry design eco-friendly homes by drawing on big data to do a lot of the hard work [8].

In a recent conference, Flux's CEO first addressed the problem by providing a simple calculation as shown in Figure III(a). Citing UN statistics of 3.3 billion people will be added to the urban population by 2050, and reasonable assumptions, she noted that we would need to build at least 500 buildings per day, every day in the next 35 years to meet that need.



(a)



(b)

FIGURE III
(A) CALCULATION OF NEW BUILDINGS REQUIRED BY 2050
(B) GOOGLE FLUX INTERFACE

To address this issue, one initiative was integrating building code and zoning regulations of a city and integrating with 3D models of buildings. As shown in Figure III, this allows public officials and developers to quickly identify opportunities for the constraints that exist on a particular parcel. Normally, this would take several days or months to process. But with the new platform this can be visualized instantaneously.

2) Sidewalk Labs
Google X also announced a new venture called Sidewalk Labs, a new and independent company is purported to focus on improving city living for everyone by developing new technologies to deal with urban issues like cost of living, transportation, and energy usage. The new company, based in New York, will be headed by Dan Doctoroff, formerly New York Deputy Mayor of Economic Development and Bloomberg CEO.

3) SiteCompli and BlockAvenue
The availability of Open Data has spawned new commercial enterprises. For example, SiteCompli uses NY's building open data to assist in building owners in tracking building permits and code regulations. BlockAvenue employs this data to provide granular economic ratings of each and every block of the US[9].

III. Methodology

Looking at the various research aforementioned, the goal of this research was to identify ways to perform Big Data analytics using open data sources in Korea. The research was focused on investigating two open data sources: 1) building permit data and 2) restaurant hygiene violations. This paper addresses the methodology and initial investigations performed on building permit data.

A. Collect building permit data and basic statistical analysis

Earlier this year, Korea started releasing public building information as Open Data (280 million data points). This constitutes datasets for over 950,000 buildings nationwide and has relevant data ranging from building permits, space usage, ownerships, to building energy data.
There are 200,000 building permit data points for the city of Seoul. The first step was to perform basic exploratory data analysis (i.e., grouping and sorting). R, and R studio was used as the base tools. The analysis at this stage will also reveal the need for a database structure and the use of a data warehouse or SQL framework.

B. Perform trend analysis for Kangnam District

The analysis was initially performed on Kangnam district, the area of Seoul with some of the highest real estate prices and where active developments take place. An initial clustering analysis of the building permits data is shown in Figure IV. Such analysis showed the evolving changes of developments over time. For example, recently the main developments are remodeling and small to medium sized housing to meet changing demographic needs. The spatial clustering also shows areas that are under developed.

C. Visualize results in GIS map using cityGML

An ancillary goal was to use OGC's cityGML, and Autodesk's Vasari which allows rapid mass modeling of buildings. Visualization of the buildings can be conducive in interpreting development activities and used to identify additional insights.

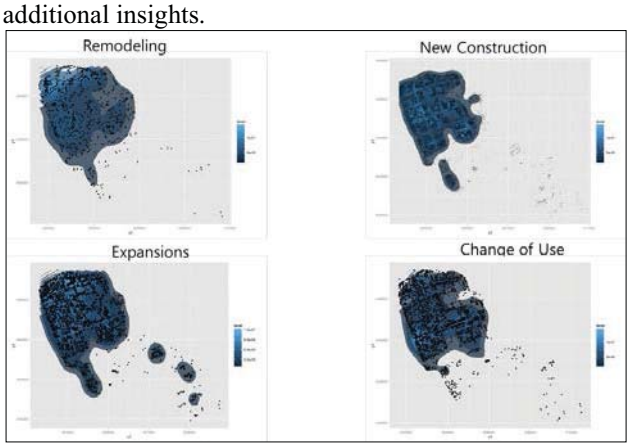


FIGURE IV
VISUALIZATION OF BUILDING PERMIT DATA OVER TIME FOR KANGNAM DISTRICT

III. Conclusion

The opening of public data and the coupled advances of data science technologies has enabled a new form of analytics for urban development and planning. Urban informatics' applicability ranges from major global cities to developing cities in need of smart solutions for critical issues. Initial research using Seoul's open data promises to provide a wealth of new opportunities to create value and insight into city improvements and ultimately to the well-being of its citizens.

References

[1] United Nations, Department of Economic and Social Affairs, Population Division (2015). World Urbanization Prospects: The 2014 Revision, (ST/ESA/SER.A/366).
[2] http://www.urbanccd.org/about-urbanccd
[3] Designing with data: Shaping our future cities. Royal Institute of British Architects. 66 Portland Place. London W1B 1AD.
[4] http://www.londonheatmap.org.uk/Mapping/
[5] http://blogs.worldbank.org/transport/what-does-big-data-have-do-owl
[6] World Bank Group Finances (2013). "DC Big Data Exploration Final Report. " World Bank Publications. http://blogs.worldbank.org/opendata/scenes-dc-big-data-dive-final-report
[7] http://www.unglobalpulse.org/
[8] http://techcrunch.com/2014/05/06/flux-the-first-startup-to-spin-out-of-google-x-nabs-8m-for-its-eco-home-building-platform/
[9] http://therealdeal.com/issues_articles/behind-big-data/