Urban Data Games: creating smart citizens for smart cities

Urban Data Games:

Creating smart citizens for smart cities

Annika Wolff, Gerd Kortuem and Jose Cavero
Department of Maths, Computing and Technology
The Open University
Milton Keynes, UK

***Abstract*—A bottom-up approach to smart cities places citizens in an active role of contributing, analysing and interpreting data in pursuit of tackling local urban challenges and building a more sustainable future city. This vision can only be realised if citizens have sufficient data literacy skills and experience of large, complex, messy, ever expanding data sets. Schools typically focus on teaching data handling skills using small, personally collected data sets obtained through scientific experimentation, leading to a gap between what is being taught and what will be needed as big data and analytics become more prevalent. This paper proposes an approach to teaching data literacy in the context of urban innovation tasks, using an idea of Urban Data Games. These are supported by a set of training data and resources that will be used in school trials for exploring the problems people have when dealing with large data and trialling novel approaches for teaching data literacy.**

Keywords—smart city, urban data, urban innovation, sustainability, data literacy

I. INTRODUCTION

Increasing quantities of data is being generated and stored from our daily lives, leading to a rise in big data analytics. Our research focuses on smart city technologies that aim to use the big urban data sets generated as people live, work and move around a city. A smart city is one in which digital technology utilises urban data to improve efficiency and sustainability [1]. Common smart city goals are better energy management, reduced water consumption and improvements to citizen mobility. Examples include facilitated taxi-sharing based on up to date location data [2], participatory noise pollution monitoring using citizens' mobile phones as sensors [3], or smart grid technologies which use fine-grained energy monitoring and prediction to improve the efficiency of electricity distribution [4]. Sources of urban data typically include: sensor measurements (e.g. air pollution, parking availability, crowds, noise); smart meter energy readings from homes and offices; up to date public transport information; crime data; social media data such as twitter data or venue check-ins. What defines these as big data is that they update frequently so that they rapidly achieve a high volume, they can show huge variation across days, seasons and years, and they are prone to error. This creates a number of both technical and conceptual difficulties for dealing with big urban data.

Key to achieving smart city goals is to adopt both a 'top-down' approach, where the technology is driven by requirements of industry or city planners and a 'bottom-up' citizen-led approach to urban innovation [5]. In this case, citizens are active in contributing ideas, collecting and analysing data and designing solutions to address local urban issues. This vision of a bottom-up smart city is reliant upon the citizens in question having sufficient data literacy skills. Research by National Numeracy[1], suggests that at least in the UK this is not the case. A recent survey revealed that 4 in 5 adults have a low level of numeracy. A key issue identified was that maths taught in school did not overlap enough with maths that could be used in later life. In fact, in the current curriculum it is clear to see that data skills are being taught, but not in an integrated fashion.

This paper poses the question how can data skills be taught to 5-18 year olds, in an integrated way that supports the use of acquired skills beyond the classroom, particularly for urban innovation and sustainability. The work is conducted in the context of the MK:Smart project[2] which is developing a data hub for storing multiple urban data sets for Milton Keynes (MK) and developing applications in the domains of energy, water, transport and citizen engagement. We have developed an approach for teaching data skills through Urban Data Games (UDG), which are playful hands-on activities with real data sets, tackling genuine urban problems such as are being addressed within MK:Smart. Games are supported by a set of learning resources for developing key skills for selecting, cleaning, interpreting, analysing and visualising big data sets. The aim is to both motivate and support participants to firstly develop and then apply key data skills in pursuit of a goal. Our data includes smart meter energy consumption, solar PV energy generation, satellite data showing solar PV potential, and other background data, such as demographic information. The UDG approach is designed to overcome some of the current barriers to teaching data skills in schools, namely lack of access to appropriate data sets, or focused activities which make use of data in a coherent context and which unify the set of data skills currently taught cross many curriculum subjects. UDGs are designed for flexible delivery to fit with classroom schedules and curriculum requirements. The UDG goal motivates teaching and learning of data skills required for competitive participation. The supporting resources can be incorporated in appropriate classroom sessions leading up to the start of the game. The approach has been designed in consultation with 5 Milton Keynes teachers from 3 different schools.

[1] http://www.nationalnumeracy.org.uk/what-the-research-says/index.html
[2] http://www.mksmart.org/

II. URBAN DATA GAMES

An Urban Data Game (UDG) is a serious game designed to motivate the learning of data skills for big, live, data sets. Based on principles of narrative, game-based learning, inquiry, collaborative learning and challenge, the following set of requirements have been identified. These specify that a UDG must have:

1. TIME FRAME: for completing all, or parts, of the task.
2. DATA. A selection of big data sets. Currently we have: smart-meter energy consumption for individual homes; satellite data from aerial surveys providing information about suitability for solar PV installation based on roof direction and roof pitch; citizen concerns; weather information; charging patterns of electric vehicles; demographic data.
3. NARRATIVE. The story that frames the task, whether this is a fictional scenario or a sequence of tasks leading to the specified ending/goal. The purpose is to provide a context against which participants can understand their roles and activities and interpret outcomes.
4. GOAL. A clearly defined set of criteria for completing a game.
5. DATA SKILLS. A clear indication of data skills needed to participate in the game. Examples include: selecting, cleaning, critiquing, contributing, combining, analysing, interpreting, and visualising data, as well as issues of privacy and security.
6. COMPETITION and JUDGING. Games are competitive and team-based. There must be a clearly defined set of criteria against which outputs are evaluated and winners chosen.

So far, two types of challenge have been identified as fitting these requirements, *Eco-puzzle* and *Appathon*. In an *Eco-puzzle*, participants are provided with multiple, partially simulated data sets where some of the data provides clues to a recent disaster. The goal is to find out what has happened in the shortest time and present a compelling justification, backed by data visualisation and analysis. Participants must query the data sets and narrow down the time frame for an event. Data will include sensor readings that show an anomaly, simulated Twitter data of some eye-witness accounts, traffic data showing travel disruption, police reports, or air pollution measurements. The goal is to apply data interpretation, analysis and visualisation skills to first identify an anomaly in one of the data sets which gives a narrative *setting* (a time and place) for the event. This is used to frame further queries across other data sets and build evidence about what occurred. Hints can be given, as there is a single, verifiable, solution but the sources of data used to evidence it can differ. Therefore, the competition may be judged on fastest or best solution, also taking into account the number of hints used.

In the Appathon, participants must design (and produce, depending on the criteria set) an App to address an urban challenge. We have created an example using data obtained from satellite survey which shows the potential for MK homes to produce solar energy. Participants must produce a visualistion of this data and identify how it might be used, e.g. as a tool for people to investigate the costs and benefits of solar PV for their own home, or to identify estates where solar PV is more or less profitable. Participants might integrate additional data sets in their app, such as energy tariff or weather data. The Appathon idea itself is not new, however unlike most Hackathon/Appathon events in which the focus is on coding skills and development of an App, the novelty in our approach is that we focus on the development of the data skills and use the goal of taking part in the Appathon as a motivator.

The resources to support the UDGs will include tutorials, tools and activities. *Tutorials* provide basic introduction to key data skills such as selecting, cleaning, critiquing, contributing, combining, analysing and visualising data. The topics are organised to reflect stages of a data-driven inquiry process. *Tools* for cleaning, analysing and visualising data are provided along with instructions for getting started. *Activities* are linked to one or more data skills, and are hands-on practical tasks, or online interactive games of short duration, for varying age and skill. They are designed to address common challenges of working with big data sets and to teach transferrable skills from the training dataset to data supplied as part of the UDG. A key focus is to trial different approaches for making large data sets more accessible to people of all ages. Examples, using our existing home-energy smart-meter data and solar PV generation, include (but are not limited to):

- creating tangible or screen-based eco-visualisations of home energy data to prompt deeper thinking about 'invisible' data
- collecting personal data in order to contextualise larger data sets, e.g. using a mobile phone as a noise sensor, or doing an energy survey of the home and school.
- Annotating/creating graphs of energy consumption and energy generation from solar PV for different times of day and curating the graphs into a coherent story, to explore how narrative principles such as *setting* can help in selecting, analysing and presenting data.

III. CONCLUSION

Urban Data Games focus the learning and application of data skills around a clearly defined task. The games are supported by a set of activities that address particular challenges associated with working with big data sets.

IV. REFERENCES

[1] P. Neirotti, A. De Marco, A. Cagliano, G. Mangano and Scorrano, F. Current trends in smart city initiatives: some stylised facts, Cities, 2010, 38, 25-36.

[2] S. Ma, Y. Zheng, and O. Wolfson. T-Share: a large-scale dynamic taxi ridesharing service, IEEE, April 2013

[3] N. Maisonneuve, M. Stevens and B. Ochab. Information polity: the international journal of government & democracy in the information age. 2010, Vol. 15 Issue 1/2, p51-71

[4] S. Karnouskos and T. de Holanda. simulation of a smart grid city with software agents. Computer Modeling and Simulation, 2009., pp.424, 429, 25-27

[5] http://www.centreforcities.org/wp-content/uploads/2014/08/14-05-29-Smart-Cities-briefing.pdf