

Urban Data in the primary classroom: bringing data literacy to the UK curriculum

Annika Wolff, Gerd Kortuem and Jose Cavero

Department of Maths, Computing and Technology

The Open University

Milton Keynes, UK

{annika.wolff, gerd.kortuem, jose.cavero}@open.ac.uk

ABSTRACT

As data becomes established as part of everyday life, the ability for the average citizen to have some level of data literacy is increasingly important. This paper describes an approach to teaching data skills in schools using real life, complex, urban data sets collected as part of a smart city project. The approach is founded on the premise that young learners have the ability to work with complex data sets if they are supported in the right way and if the tasks are grounded in a real life context. Narrative principles are used to frame the task, to assist interpretation and tell stories from data and to structure queries of datasets. An inquiry-based methodology organises the activities. This paper describes the initial trial in a UK primary school in which twelve students aged 9-10 years learnt about home energy consumption and the generation of solar energy from home solar PV, by interpreting existing visualisations of smart meter data and data obtained from aerial survey. Additional trials are scheduled with older learners which will evaluate learners on more challenging data handling tasks. The trials are informing the development of the Urban Data School, a web-based platform designed to support teaching data skills in schools in order to improve data literacy among school leavers.

Categories and Subject Descriptors

J.2 [Computer Applications]: Physical Sciences and Engineering – *mathematics and statistics*

General Terms

Design, Experimentation, Human Factors

Keywords

Data literacy, curriculum, smart city, urban data, narrative, inquiry

1. INTRODUCTION

The term data literacy is generally used to refer to a set of abilities relating to the use of data [11]. These include the ability to identify sources of data, to ask questions of data, use appropriate tools for statistical analysis and visualisation of data, to use data as evidence to support an argument and to be able to understand and interpret the data that others use as evidence. It includes the ability to work with large data sets and to think critically about data, as it is encountered in every day life. What distinguishes data literacy from statistical literacy is the embedding of data in a context through which the data is interpreted, the focus is on the task as a whole rather than the application of statistical methods [18].

Data literacy is becoming a big topic. The Employer Insights: skills survey 2015 found that 27% of the biggest employers (> 250 employees) now use data analytics to support their work [16]. But businesses also report that they cannot hire the data scientists they need to analyse their growing datasets and ensure their competitiveness [3]. The House of Lords Select Committee on Digital Skills released a report in February 2015 [8] in which they suggest that digital skills, in particular data skills necessary to protect oneself online, are now ‘necessary life skills’. They further argue the importance of ensuring that the majority of the population achieve an appropriate level of digital literacy to be full participants of the increasingly digital future. In a similar vein, the emerging topic of human-data interaction (HDI) [7] investigates how the combination of personal data with other sources and lack of transparency in the applied algorithms affects individuals ability to understand how their data is used, or misused. In addition to data owned and used by companies, Open Data initiatives are making data more freely available. A proposed use for open data is to drive bottom-up citizen-led smart city innovation, as well as provide additional resources for commercial enterprise [12][13], though it is hard to find evidence of this data being used outside organised hack events, by participants with above-average data skills. Taken as a whole, this points to a need to raise the level of data literacy within society.

In this paper we describe an approach to improving data literacy in young learners by supporting them in hands on activities with complex, urban, data sets. We identify some of the barriers to teaching data literacy using these types of data sets in the classroom and some approaches to overcome these problems. We describe our initial trial in which primary school children interpreted data visualisation related to energy consumption and generation through solar PV. We describe the next phase of trials in which secondary students will undertake more complex data analysis of the same data sets. We propose that it is important to give learners the opportunities to gain practical experience of working with large data sets from a young age. We suggest that these types of tasks should be integrated into the standard curriculum in order to raise standards of data literacy amongst school-leavers, to give them more options within the future job market and to become more participating citizens in smart city innovation.

The work described in this paper has been conducted as part of the MK:Smart project¹. MK:Smart is developing smart technologies for the City of Milton Keynes (MK), UK, in three

¹ <http://www.mksmart.org/>

key areas: energy, water and transport. Central to MK:Smart is a data hub which is aggregating both specific project-related data sets and other open data sets. The data hub is then available as a resource for citizens to develop innovation projects, using data as evidence, and as an educational resource for use by schools. To this end we are developing the Urban Data School (UDS), a web-based teaching and learning platform through which teachers can access and contribute data and resources for teaching data skills in schools. This platform will facilitate the scaling up of our approach.

2. RELATED WORK

This section describes learning approaches that have informed the development of our approach to using urban data sets as a resource for teaching data skills in schools.

2.1 Data analysis as inquiry process

Inquiry-based learning is a constructivist approach in which the learner is supported to take an active role in designing and running their own experiments [6]. An inquiry-based task generally follows a cycle of asking questions, investigating through data collection, analysing/organising the data, discussing and answering questions and reflection (Figure 1). In a scientific inquiry, data is collected as evidence to test a hypothesis. A historical inquiry uses primary (first hand accounts) and secondary sources of evidence to create historical narratives.

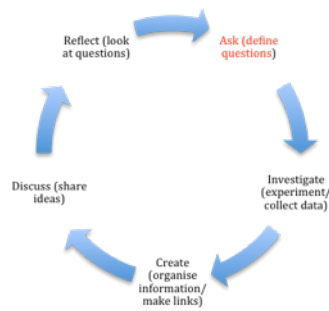


Figure 1. A typical inquiry cycle

PPDAC applies an inquiry-based approach to statistical analysis from data [19]. PPDAC stands for **P**roblem, **P**lan, **D**ata collection, **A**nalysis and **C**onclusion (Figure 2). This approach is taught in schools in New Zealand, where statistics is emphasised as a subject in its own right and is taught alongside (not as part of) mathematics. By contrast, in the UK, statistics is mainly taught in an abstract way as part of the maths curriculum. Whilst PPDAC is designed to start with a problem specification that leads to some data collection, it is possible to find examples of lesson plans that propose to start with an existing dataset and work backwards to understand what problems could be answered with the data, for example to look at noise mapping from UK noise data². Posing questions from data is a key part of the PPDAC data inquiry cycle [2]. This approach emphasises the importance of context within statistical inquiry, for generating good questions to ask from the data and for interpreting the results of statistical analysis [1].

A number of toolkits or task frameworks have been developed to support a data inquiry approach. Data Toolkit³ provides a framework, tutorials and tools aimed at arts

organisations who want to work with young people in developing data-driven arts projects. Completed projects are loaded back to the site following a standard template, to be used as inspiration/case studies for future projects. School of data⁴ follows a similar approach to the Data Toolkit, but gives a stronger narrative by framing the task as a 'data expedition' and outlines how a 'data guide' can structure the inquiry as a collaborative effort, assigning different roles such as 'story-teller' or 'scout'. Participants might choose a role that will develop their skills in a specific area of the data inquiry. The intended audience for the School of Data are journalists, or citizens who want to learn how to use data more effectively.



Figure 2. The PPDAC cycle

2.2 Narrative principles

Stories are a natural means through which people make sense of the world [14]. Narrative has been shown to be particularly important to young learners in assisting recall [20]. Narrative provides coherence through the people and setting (time and place) of the story and the use of theme [5][17].

Data visualisations are often used to present data in a coherent way that can support the telling of a story [15]. This type of storytelling with data focuses on using narrative principles to construct a story to communicate findings, rather than as a tool to guide interpretation and analysis from data.

3. AN APPROACH TO USING URBAN DATA IN THE CLASSROOM

The following sections describe how we apply narrative principles and adopt an approach based on the PPDAC inquiry cycle to support learners working with urban data sets.

3.1 Urban data

The MK:Smart project is aggregating a large amount of urban data. We define urban data as data that informs about environmental, social and economic aspects of urban life. It is typically generated by the activities of people living in a city. The data relates to a variety of topics, such as air and water pollution, energy and water consumption, crime, public and private transport, or car parking to give some examples. Some of the data is available in the form of live data streams that are real-time indicators of the state of the city system (such as traffic flow). Such data streams can generate large data volumes that show variations across days, seasons and years, and – more often than not – are affected by data quality issues. They can cover a large geographic area and/or time-span. Urban data may be collected through sensors, smart meters, satellite imagery or mobile phones, or derived from surveys and questionnaires.

² <http://www.spatialanalysisonline.com/noise.pdf>

³ <http://datatoolkit.org.uk>

⁴ <http://schoolofdata.org>

3.2 Initial design and theoretical approach

Schools have typically focused on teaching data skills from smaller, often personally collected, data [9]. Since larger, more complex data (such as urban data) has not typically been available as a teaching resource, there is not yet a clear research base to assess the best approach to using this data as part of classroom teaching. Larger data sets can present unique challenges in terms of selecting, cleaning, critiquing, interpreting, visualising, storage and processing compared to smaller, less complex datasets. It is not clear whether skills learnt against smaller, personal, data sets can be easily extended to apply to more complex urban data.

To develop our approach for teaching data skills with urban data we have drawn upon existing learning methodologies to inform an initial design phase. When designing tasks with urban data we use *narrative principles* to:

- provide a strong narrative to frame the task context.
- support data interpretation to tell stories about and across one or more data visualisations.
- provide assistance for the structuring of queries to large data sets by thinking about *setting*, *theme* and *actors*.

To support this, we provide students real-life data sets to work from and create activities within which they undertake genuine research as a way of providing context, rather than learning data analysis as a set of abstract techniques in a ‘dummy’ task context.

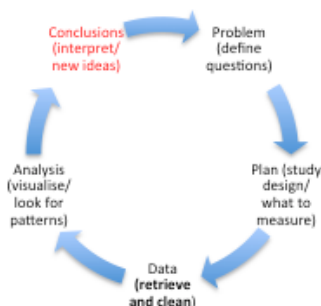


Figure 3. Modified PPDAC cycle, starting with the conclusions and using existing data.

We structure tasks as an *inquiry process* similar to PPDAC, but with some modifications to take into account specific issues related to using existing large data sets rather than small, personally collected data. In our case, the starting point for the inquiry is the interpretation of an existing visualisation. This acts as a seed for defining a new problem (figure 3). We use the idea of ‘telling stories’ from the data to help interpretation, i.e. by encouraging learners to identify the *setting* (when and where) of the data, the main *theme* of the visualisation, which might be a changing variable and the *actors*, which are responsible for affecting the variable through their actions. The actor may or may not be a person, depending on the dataset.

This approach is designed to overcome what we foresee as a ‘cold start’ problem, where learners could find it difficult to define questions that can be answered from a large data set without first forming some view of the data by interpreting existing visualisations. This view is supported by the teachers who assisted in development of the approach (see section 3.3.1).

Learners then continue to formulate questions around this same data set. In this case, the questions are used to structure queries that will retrieve some part of the large data set that may then be used as evidence in answering the inquiry question.

One possible problem with urban data in the classroom is that students might find it hard to understand how the data was collected, which can make it harder to interpret [10]. We hypothesise that learning from a large, abstract data set might be supported through a personal data collection task that is related to the original data inquiry. Essentially, students complete one or more cycles of an inquiry based around an existing large data set, then undertake a second phase where they conduct an alternative but related inquiry from their personally collected data. This is more traditionally the PPDAC scenario as shown in Figure 2. The purpose is to consolidate the learning of data skills and also to provide additional narrative and context for understanding the larger, more abstract, data set.

3.3 Developing lesson resources

3.3.1 Teacher engagement

We adopt a participatory design approach in which both teachers and students inform the development of methods and materials for teaching data literacy in schools using urban data. Currently we have engaged with teachers and students from 3 secondary and 1 primary school in Milton Keynes, with a focus on supporting learners aged between 8-18 years. Engagement has occurred through a series of detailed face-to-face meetings and an iteration of materials and ideas. Teachers were recruited through project partners of MK:Smart and through previous projects at the Open University.

3.3.2 Selection of data

The data sets that we are using have been collected as part of MK:Smart project activities. Appropriate data for schools have been identified through the following process. All data sets were initially categorised by theme. Next, Priority was given to data that had already been used within some well-defined smart city research projects. Finally, two data sets were chosen which were both related to the topic of energy. One data set is time-stamped smart meter energy data, detailing home energy consumption of both the whole house and individual appliances and also solar PV generation for approximately 70 houses in Milton Keynes over a 2 year period. This data set has previously been used to explore the gap between home solar generation and energy use and the potential for demand shifting to optimise use of locally produced energy [4]. A second dataset contains data obtained from aerial survey that can be used to estimate the solar energy yield from existing or potential solar panels on the roofs of houses in Milton Keynes, using a pixel-based analysis to identify pitch, orientation and surface area. The examples we will use in this paper are based on real data, but for the purposes of publication have been completely anonymised.

3.3.3 Creating lesson plans

A set of activities have been devised from the data sets for use in classrooms. The first activity uses the smart meter data set to pose questions around how and when energy is consumed in an average household and when solar panels on a roof are producing energy. The second activity uses the aerial collected data to pose questions related to whether or not properties are suitable for having solar panels. These activities are being tailored to different age-groups and teacher preferences and adapted according to feedback from trials. The overall session format consists of

1) introduction to the topic 2) One or more data inquiries following the methodology described above 3) Conclusions. The following section describes the process through which these activities were used in a primary school classroom.

4. CLASSROOM TRIAL

The main aim of the trial was to elicit initial feedback on the materials and to gain an understanding of how teachers would use the resources in a real classroom setting. To allow the teacher full flexibility in choosing how to use the content, it was decided that evaluation should be through a mixture of classroom observation followed by more structured feedback by teachers and students at the end of the session. The focus of the evaluation would be related to a) how well students were able to interpret visualisations of complex data sets b) how easily the materials could be incorporated into regular classroom teaching c) the students engagement with and learning about the chosen topic ‘energy’.

The classroom trial was conducted over two normal class session in two consecutive weeks, with twelve Year 5 students (aged 10-11). These students were in a maths group and were described by the teacher as being more used to being challenged with novel materials and approaches.

The format and teaching materials were developed through close collaboration between the researcher and the teacher in the weeks leading up to the sessions being conducted. This consisted of a series of face-to-face meetings and detailed discussion of all materials that were to be used in the classroom, in addition to some email exchanges. Based on teacher feedback it was decided that primary students would focus on interpretation of existing data, rather than conducting their own data analysis. In terms of PPDAC, this meant that the data retrieval step would involve the retrieval of ‘canned’ visualisations or tables, however students would still be required to clean and find patterns in the data to answer questions.

The classroom sessions were led jointly by the researcher and the teacher, with one further researcher joining in the role of formal observer. The observation strategy was discussed prior to the session. The role of the observer was to record student responses to questions during the ‘whole class’, noting observations on students understanding of the questions and whether questions answered were prompted or unprompted. At some points, students worked in three smaller groups. In this case, each researcher acted as observer for one group following the same observation protocols. The teacher was tasked with supporting and observing the third group.

4.1 Session 1

4.1.1 Establishing the context: introduction to energy

To introduce the topic of energy students were taken through a set of prepared slides. This was an interactive session, intended to elicit feedback from students about their knowledge of things related to how energy is produced, consumed and measured. Students were asked to name things that consumed electricity. Students worked firstly in pairs (6 groups of 2) and wrote their ideas down on a small whiteboard. Some of these whiteboards were photographed (figure 4).

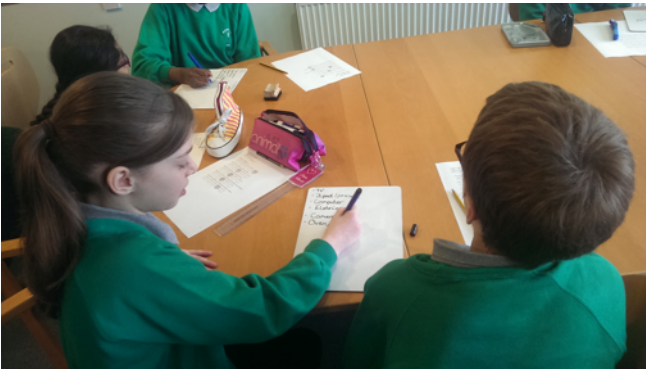


Figure 4. Naming items that consume energy

Some of these ideas were then put forward to the group and discussed. Students identified a wide range of items that consumed electricity both inside and outside the home. Examples included ‘toaster’ ‘robot’ ‘traffic lights’. The full list of items either written down or mentioned can be seen in table 1. Whilst there were a couple of odd responses (e.g. solar panels and satellites) overall students demonstrated a very good understanding of electricity consumption.

Table 1. A list of items identified by the class that consume energy

• Light bulb	• Mobiles
• Kettle	• Tablet
• Microwave	• Ipad
• Oven	• Video games
• Dishwasher	• Wii
• Toaster	• Camera
• Fridge	• Solar panels
• Freezer	• Electronic devices
• Washing machine	• Satellites
• Iron	• Robots
• Electric vehicles	• Hoover
• Traffic lights	• Car
• Computer	• Radio
• Laptop	• Pylon

Next, students were asked to name some things that generate electricity and to discuss whether they were green or not. Students answered by putting their hands up. The first answer was ‘solar panels’ and the second answer was ‘pylons’. It was explained by the teacher that pylons are used to transport energy. After being asked again, one of the students mentioned that burning coal produces energy. Prompted towards the problems of burning coal one student identified that coal was a resource that could run out leading to problems in the future. The researcher discussed briefly about the impact on the environment from burning fossil fuels. Next they were asked ‘how can we generate electricity in a more green way?’ The first answer was that we can use ‘the sun, because it is so hot’. After being asked again they mentioned ‘wind turbines’ and ‘water’ as ways of generating electricity.

Overall, it was established that students had a good understanding of the relevance of green energy, but needed some prompting to draw out their knowledge.

Next there was a brief introduction led by the researcher about measuring electricity consumption through a smart meter.

4.1.2 Group work

Students worked in three groups of 4 students on different tasks related to energy consumption. The tasks were designed so that students would apply similar data skills within each group but would learn about a different aspect of home energy consumption or generation that they could then share amongst each other, to become teachers as well as learners of new knowledge. It was explained to the students that each group would be given a set of graphs and a question to answer. They should first look at just *one* graph to understand what it showed and to try and tell a story about it, then they should look at all the materials and try to answer the questions as a group. Each group would have a chance to present their findings back to everyone at the end of the session. The following sections describe the activities of each group in detail and how successful they were in answering their questions.

4.1.2.1 Group 1

Group 1 were asked to look at home energy consumption on a daily basis. The question they had to answer was:

When is a) most energy consumed b) least energy consumed in an average household.

They were provided with graphs of daily consumption for 2 different houses, identified to them as house 4 and house 29. The first graph that they looked at is shown in figure 5.

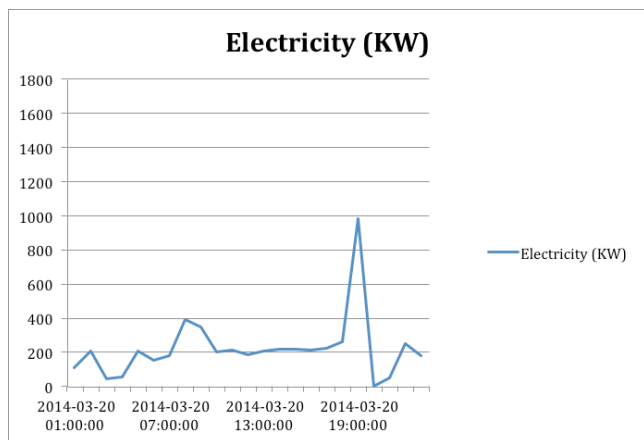


Figure 5. Whole house energy consumption in 1 day

The observations of this group work are as follows:

From the first graph (figure 5) students quickly identified that the graph was showing one day, from morning to evening and that the Y-axis represented electricity use. Students in this group offered, without having to be prompted, the view that the evening spike might be caused by 'people coming home from school at 7 or people working'. Students were prompted to explain what might be causing some electricity to still be consumed during the night. Students thought that it could be people staying up in the night watching television. Again students were prompted to 'think what appliances are in the home that don't get turned off'. Following this prompt, one student suggested that fridges would

be left on overnight and would use electricity. Students were occasionally writing onto their whiteboard as they worked. Figure 6 shows one student who had drawn (and is seen rubbing out) the graph pattern and writing their theory about the pattern.



Figure 6. Interpreting graphs of energy consumption

Students were prompted to look at further houses to see if the same pattern was happening. Students were given access to 8 days of consumption from House 4 and 8 days from House 29. Students noticed that the pattern was normally the same but when looking at one that was different one student offered 'House 29 is different. It could be Saturday. There are 8 days here' the implication being that because there were eight days then at least one of the days could be a weekend.

Group 1 wrote down their ideas about how to answer the questions onto their whiteboards for presenting back to the group. These are shown in figure 7.

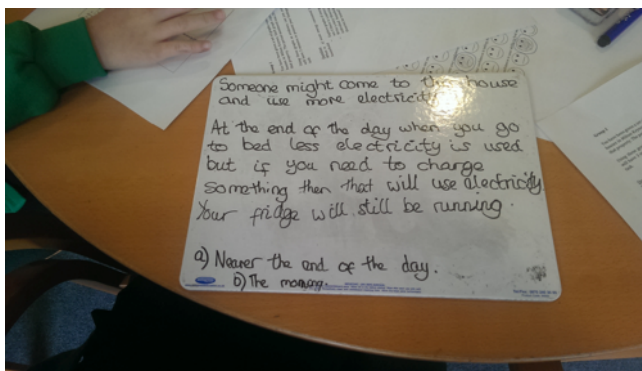
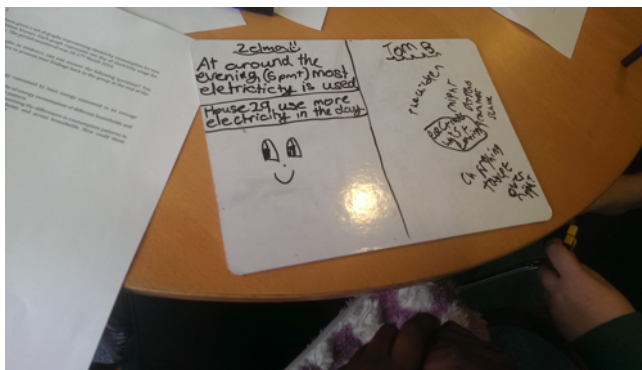


Figure 7. Creating theories of energy consumption

Group 1 presented their findings back to everyone. One of their graphs was shown on the projector. Students explained in this presentation that more electricity might be used overnight because people were 'charging overnight'. They proposed that 'not so much [electricity] is used in the morning or night because people are asleep and it is light in the day'. The teacher prompted

the rest of the class to think about what the graph showed. Students suggested that the electricity used at night might be because of a fridge being on or because people were scared of the dark and left the light on.

Overall, group one were observed to have a very good understanding of the graphs, were able to interpret them correctly with little prompting, form ideas about what was causing different electricity consumption at different times of day, again without prompting. However, they needed prompting to start looking at patterns across multiple graphs. The main suggestion of the researcher from observing group 1 is to only show graphs of consumption from one house, rather than two. It is further proposed that interpreting one graph to begin with was a successful approach and that in keeping with this approach it would be sensible to prompt incrementally adding more graphs to build up understanding. For example, to interpret and tell a story about one graph, present a series of weekday graphs for the learner to discover a pattern of typical weekday energy consumption. When this pattern is established a graph of 'weekend' activity can be shown.

4.1.2.2 Group 2

Students were given graphs of energy use by both a washing machine and a fridge, in firstly a short time (hourly) and secondly a longer time (3 months). Group 2 were asked the question:

Do appliances use the same energy as each other? As themselves? Which appliance uses most energy overall?

The observations of this group work are as follows. At the beginning of the activity the researcher noted that the students were looking to the graphics and questions but without making any comment. Then they started to look at the "hourly time scale" washing machine graphic and the researcher prompted them towards what happen when you switch on the washing machine, particularly discussing if the level of noise is always the same while the machine is working. The researcher felt that they easily understood that it varies during washing machine use and that this variation is related to the level of energy consumed. After this they have guessed what could have happened in each graphic. The researcher felt they have understood the fridge graphic very well, correctly interpreting why sometimes it is possible to find some peaks in the energy consumption. They have referred to the fridge in terms of using or not using it. They have said that the peaks are caused "when you are using the fridge" referring to when you are opening and closing it, but they were not particularly explicit. They also mentioned also that the base level occurs for example at night 'when you are not using it'. Students were observed to note that the washing machine and the fridge have different "base energy" levels. They then moved to "daily time scale" graphics, talking again about the base energy consumption level and explaining that the fridge is always used while the washing machine is not.

The researcher asked why there are some fridge and washing machine graphics that look so different to each other and the group correctly answered that is because of the difference of the time scale, being hourly in the first graphic and daily in the second. The only real difficulty that the researcher noted was when talking about total energy consumption, particularly in 2 aspects: The students understood what the Y-axis represented (energy consumption) but they have not checked the values when comparing appliances and they have just assumed that all the graphics are in the same scale. The researcher had the impression

that they don't completely understand that energy consumption corresponds to the area under the graphic.

In presenting back to the group, the student proposed that fridges use the most energy overall. They requested that the graph of 3 months of fridge electricity consumption was shown to the class and contrasted against the washing machine use in the same period. They suggested that in the washing machine someone turns them on and off and the fridge is on all the time.

4.1.2.3 Group 3

Group 3 were given graphs showing the generation of solar energy in 5-day periods in both June and December, for three different houses. The first graph that the students interpreted is shown in figure 7. Group 3 were asked:

What factors affect the amount of energy generated from solar PV?

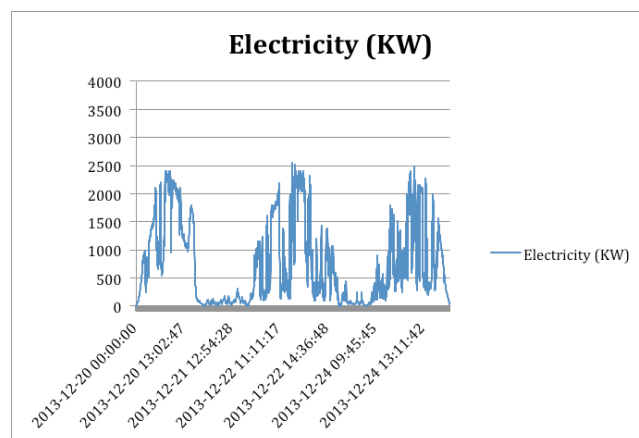


Figure 7. Graph showing generation of solar energy for 5 days in December.

Group 3 were supported by the teacher, who had not been briefed on the observation protocols. For group 3 we have the verbatim presentation back to the class, as below:

'We had three different houses and house 15 measured 5 days of generating electricity in June and 5 days in December. We found out that on house 15 it generates the most electricity than house 29 on the same days by just a couple of thousands.'

The teacher then prompted to ask what could explain this, the student replied:

'The solar panels near their house might be huge or they might have different ones but in House 29 they might have smaller ones and it might not be that good.'

The teacher prompted what might explain the difference between June and December, the student replied:

'In June because you got more hours of the day there's more sun - more chance of sun, and the solar panels need sun to create electricity and in winter its normally rainy and cloudy and snowy so not normally so much sun and not too many hours where the sun is shining.'

4.1.3 Closing

To conclude the inquiries, students were shown a graph highlighting the gap between energy generation from solar PV and consumption (figure 8). Students were led by the teacher in a discussion related to how most households consume energy in the evening whereas solar energy is generated in the daytime.

Figure 1 is a line graph comparing Consumption (red line) and Solar generation (green area) over a 24-hour period. The Y-axis represents Generation/Consumption in MW, ranging from 0 to 5,500. The X-axis represents time in hours from 00:00 to 23:00. Solar generation peaks around 12:00 at approximately 3,500 MW. Consumption peaks around 18:00 at approximately 4,500 MW.

Students were given homework to ‘Be a Smart Meter’ and make a note of their energy consumption habits between that session and the next.

4.2.1 Introduction to the data

4.2.2 Group work

Year	1970-1979	1980-1989	1990-1999	2000-2009	2010-2019	2020-2029	2030-2039	2040-2049	2050-2059	2060-2069	2070-2079	2080-2089	2090-2099	2100-2100		
Age	14.7370250	19.0751	13.7525	13.6918	2.90	1.84	2.615	14.39	2060	234.03	9421.37	847.07	8218.88	9068.88	8.7	12.78
Age	17.18478	24.9679	16.943	16.9187	2.90	1.84	2.615	14.39	2060	234.03	9421.37	847.07	8218.88	9068.88	8.7	12.78
Age	17.000790	24.9679	16.943	16.9187	2.90	2.07	2.011	14.39	4140	2040.0	6672.7	1009.29	162.37	11879.26	0.86	15.5
Age	14.7370250	19.0751	13.7525	13.6918	2.90	1.84	2.615	14.39	2060	234.03	9421.37	847.07	8218.88	9068.88	8.7	12.78
Age	14.5175	13.2089	24.0193	24.0193	2.90	1.84	2.615	14.39	2060	234.03	9421.37	847.07	8218.88	9068.88	8.7	12.78
Age	14.5175	13.2089	24.0193	24.0193	2.90	1.84	2.615	14.39	2060	234.03	9421.37	847.07	8218.88	9068.88	8.7	12.78
Age	22.981113	24.0193	24.0193	24.0193	2.90	3.45	2.000	14.39	4140	2040.0	6672.7	1009.29	162.37	11879.26	0.86	15.5
Age	22.981113	24.0193	24.0193	24.0193	2.90	3.45	2.000	14.39	4140	2040.0	6672.7	1009.29	162.37	11879.26	0.86	15.5
Age	22.981113	24.0193	24.0193	24.0193	2.90	3.45	2.000	14.39	4140	2040.0	6672.7	1009.29	162.37	11879.26	0.86	15.5
Age	22.981113	24.0193	24.0193	24.0193	2.90	3.45	2.000	14.39	4140	2040.0	6672.7	1009.29	162.37	11879.26	0.86	15.5
Age	22.981113	24.0193	24.0193	24.0193	2.90	3.45	2.000	14.39	4140	2040.0	6672.7	1009.29	162.37	11879.26	0.86	15.5
Age	22.981113	24.0193	24.0193	24.0193	2.90	3.45	2.000	14.39	4140	2040.0	6672.7	1009.29	162.37	11879.26	0.86	15.5
Age	22.981113	24.0193	24.0193	24.0193	2.90	3.45	2.000	14.39	4140	2040.0	6672.7	1009.29	162.37	11879.26	0.86	15.5
Age	22.981113	24.0193	24.0193	24.0193	2.90	3.45	2.000	14.39	4140	2040.0	6672.7	1009.29	162.37	11879.26	0.86	15.5
Age	22.981113	24.0193	24.0193	24.0193	2.90	3.45	2.000	14.39	4140	2040.0	6672.7	1009.29	162.37	11879.26	0.86	15.5
Age	22.981113	24.0193	24.0193	24.0193	2.90	3.45	2.000	14.39	4140	2040.0	6672.7	1009.29	162.37	11879.26	0.86	15.5
Age	22.981113	24.0193	24.0193	24.0193	2.90	3.45	2.000	14.39	4140	2040.0	6672.7	1009.29	162.37	11879.26	0.86	15.5
Age	22.981113	24.0193	24.0193	24.0193	2.90	3.45	2.000	14.39	4140	2040.0	6672.7	1009.29	162.37	11879.26	0.86	15.5
Age	22.981113	24.0193	24.0193	24.0193	2.90	3.45	2.000	14.39	4140	2040.0	6672.7	1009.29	162.37	11879.26	0.86	15.5
Age	22.981113	24.0193	24.0193	24.0193	2.90	3.45	2.000	14.39	4140	2040.0	6672.7	1009.29	162.37	11879.26	0.86	15.5
Age	22.981113	24.0193	24.0193	24.0193	2.90	3.45	2.000	14.39	4140	2040.0	6672.7	1009.29	162.37	11879.26	0.86	15

In one map and table there was data for existing panels and in the next there was data for potential installations (Figure 10). Students were asked to use the data to discover the relationship between roof area and yield, to understand how other factors such as pitch and direction influenced the yield and then to consider the different types of information they could get from the two data sets (aerial image and data table). Next, students were given the data for their own school and were asked to consider whether it was cost effective for their school to have solar panels. As in the previous session, a teacher and researcher visited the groups to help with any difficulties and to observe their progress.

pictures' whereas another offered 'I liked the table it's easier to see the data'.



When prompted to identify something that could be seen in the map that didn't appear in the table, one student (wrongly) suggested the pitch of the roof. Another student then correctly identified that it was possible to see trees in the aerial image and the trees might get in the way of the solar panels. Students did not write down the answers to questions in this session but were asked to summarise their feedback, therefore there is no quantitative data relating to how well students answered the question for identifying the best and worst houses on the estate for producing solar panels. However, the observations of the researcher was that students had correctly identified good and bad houses from the map, and further that students who preferred looking at the data in the table could easily find the highest and lowest figures for yield and understood that these were used to answer the question.

'We found out that if the house roof are slanted then they have the most chance of getting the most electricity but if they're flat they probably get a bit but not as much.'

'If you had a 4 sided roof it would be harder to put solar panels on because sometimes the sun doesn't come from that side so not all roofs are ok for putting panels on.'

‘It’s best if the solar panels are facing south because that’s the direction of the sun in the day.’

‘Even if you buy these really big expensive solar panels it might not make much of a difference it might be a waste of money.’

Students were then prompted to look at the data for their own school in the potential installations map (Figure 10). This required students to use additional columns in the table relating to the predicted cost of panels and estimated savings over 1 or 20 years. Students required some prompting to start looking at this additional data, and also needed an explanation of what the new columns meant. This information had been given in the form of a key for all tables, so students needed to be reminded of this. Students had 15 minutes to look at this data and then were again asked to discuss whether the school should get solar panels or not. The verbatim responses were:

‘I think we should but only in the classroom and foundation (*referring to specific buildings/roofs in the data) because the other ones are kind of useless because there are so many children in the school not at their house using electronics.’

‘I think we shouldn’t consider because it would be really a lot of money it would be over 10,000 just for one.’

‘We can get a lot of energy because the roof is bigger so we can fit a lot of solar panels on.’

‘Another good thing because the benefits of installing them would be good because we could produce more electricity in summer because days are longer.’

‘Even though it would cost a lot of money if we put several solar panels on our school it might be worth it because of the stuff we’re getting back.’

Overall, students demonstrated a very good understanding of the essential purpose of the dataset and were able to use the data to answer the questions.

The students had also been given homework to ‘be a smart meter’ and had been keeping data diaries during the week between sessions where they recorded what appliances were being used and how long for. Students were asked to note when they spotted something that was on, but not being used. They were also asked to keep an ideas log of ways to save energy and water. Only 7 students returned their data diaries. Figure 11 shows a word cloud of the most recorded appliances: According to the data, the top 3 appliances that were identified as being on but not used were:

- TV
- lights
- computing devices (laptop, apple computer....)

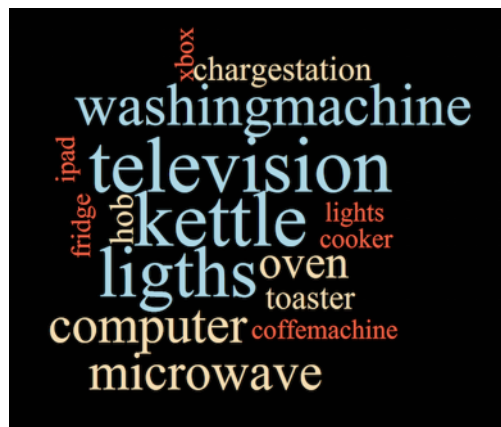


Figure 11. Word cloud of appliances that students recorded energy use for.

Three students submitted ideas through their ideas log. Overall, these demonstrated a good awareness of what is needed to save water and energy. The most popular suggestions were:

- turn lights and tv off when you leave the room
- turn off tap when not being used
- shower instead of bath
- re-use water
- turn chargers off when not in use

4.3 Post session feedback

The teacher provided their own feedback after the session by answering a semi-structured questionnaire supplied by the researcher. The teacher was also asked to elicit feedback from the class in whichever manner they felt would get the most honest response, this was to avoid bias in responses given directly from students to the ‘unknown’ researcher. 11 of the 12 students were available to give feedback. The following is a summary of this feedback.

Students were asked what they found easy/hard. Responses can be seen in Table 2.

Table 2. Post session student feedback.

	Easy	Hard	No response
Coming up with ideas and suggestions about what things meant and showed	5	1	5
Understanding the graphs	6	3	2
The homework	3	3	5

In addition, 7 students thought that the sessions were ‘well thought out and useful / well organized’. 5 said the sessions were ‘good, but I would have liked another session, so we could have gone a bit more slowly and/or learn a bit more’. 2 students responded that some of the materials were hard to understand.

When asked what key points they had taken away, 5 students responded that ‘solar panels are really good, especially if your roof gets lots of sun’ and 1 student replied that ‘some homes could generate a lot of energy but don’t’. 2 students said they would ‘try not to waste so much energy’ and 2 more said that ‘saving energy would help a lot’.

Students were asked ‘How has it changed your understanding of energy use?’ They replied:

- People use lots more energy in the evening
- I realise that the TV uses lots of energy and it could run out quickly
- I realise that saving electricity is important
- If you leave the TV on without watching it then it wastes energy

When asked if the sessions had changed how students used energy 2 responded yes, 5 responded a bit and 4 said no, it had not changed. When asked how their behavior had changed students replied:

- I switch things off more now
- I don’t charge my iPad for as long now
- I turn off chargers and lights when my family leave them on
- I didn’t realise how I was wasting energy, so now I switch things off more when I don’t need them
- Sometimes I turn things off now when people aren’t using them
- When my phone is charged I turn the socket off.

The teacher feedback was based on their own knowledge and experience of the students who participated in the sessions. The teacher was asked ‘How much do you think the students enjoyed the sessions?’ on a scale of 1 (not at all) to 5 (very much). The teacher responded between 4 and 5. When asked ‘what do you think students found easiest?’ the teacher replied ‘understanding

the graphs and the differences they showed; making reasonable suggestions about what the graphs showed – or not’. When asked ‘what do you think students found hardest?’ the teacher responded ‘A couple found it all quite challenging, and a few found it hard that the graphs were so different from what they are used to seeing. I think putting all this in two lessons was hard for some.’ However, the teacher felt overall that the materials were pitched appropriately for the age-group, with the proviso that less able Year 5 students might need to work in groups with at least one more able student to help them along. The teacher suggested that one additional session would have been helpful to allow students to learn at a slower pace. The teacher expressed interest in having access to the raw data if it were available through the Urban Data School, but felt that they would probably still use paper copies of visualisations for analysis and only have digital representations when the whole class was looking at something. The teachers final comment was that the session was ‘Interesting and useful, and I was pleased with the way the pupils coped with some challenging ideas and concepts, many of which are well beyond what they usually deal with.’

4.4 Teacher feedback

In the work so far and through discussions with teachers we have received the following feedback:

- Nearly all teachers mention that they are interested in using the Milton Keynes data as it has a local angle and provides more interest for students.
- Teachers are very keen to use the data as much as possible ‘as is’, not to ‘dumb down’ the teaching.
- Teachers think that using real data will help them to answer the common question of students ‘why do I need to know this?’

However, through these discussions we have also identified a number of potential barriers to using large data sets in the classroom. These include:

- Lack of teacher confidence/expertise in using big data sets. Teachers are keen to use the resources but prefer to have the researcher in the classroom as ‘back up’ rather than use resources unsupported.
- Even with researcher assistance, teachers show a strong preference towards doing only data interpretation of existing visualisations rather than data analysis. Issues included the need to organise computer access in subjects where they typically weren’t used (e.g. in a science or geography lesson), the need to procure software, lack of expertise in using the software, general preference for paper-based rather than computer-based teaching.

5. DISCUSSION

Twelve year 5 students were tasked with answering questions using visualisations of complex data sets as evidence. These students demonstrated that they were capable of achieving the tasks presented to them, with some level of help that would be normally expected in a classroom. Completing the tasks contributed to a developing understanding of the chosen domain of energy and prompted students in many cases to reflect on their own behaviour. Starting the PPDAC inquiry cycle from the conclusions did not appear to be a disadvantage. Students demonstrated that they were able to make valid interpretations and tell nice stories from the relatively sparse information given in the

visualisations themselves, despite not having gone through the earlier stages of planning, data collection and data analysis. This could be attributed to the use of a strong narrative theme of ‘home energy consumption and generation’ which they could use to provide context to the task and which related all of the different data sets used. Students demonstrated that they were able to tell stories about the data and to think in terms of the time and physical location of the data, e.g. talking about activities that take place in a home at certain times of day.

The overall approach has been designed so that students must demonstrate some pre-defined set of domain-independent data skills as they complete a task in a specific domain. The study described in this paper focused on only a very small aspect of data literacy - that of interpreting visualisations of complex data sets - in the context of a task focused on increasing energy literacy. It should be noted that the data and visualisations used were not modified and as the teacher identified, these are not the types of materials or concepts the students normally use. Therefore, the outcome of the trial might be taken as an encouraging indication of the capabilities of students of this age with respect to this specific data skill. Whilst it should be noted that these students were considered the more capable within their year group, which could have implications for the generalizability of these results with respect to primary students, the teacher was confident that the same task would work in the same year group using mixed-ability groups.

In order to extend the approach with respect to increasing data literacy, it will be necessary to support the learning of additional data skills, particularly those that are problematic when working with larger datasets. These include formulating queries for data selection, cleaning and visualising large data. In this regard, it might be expected that if primary student are capable of interpreting the data then older secondary school students, with similar support, would be able to go one step further in conducting their own queries and creating visualisations.

6. FUTURE WORK

The next step will be to evaluate the same materials in a classroom session with secondary students. In this case, students will be given the identical task of storytelling from a single visualisation. However, these students will then be tasked with using this story as a basis for posing further questions. For example, a student interpreting a graph of home energy consumption might ask questions about whether the pattern of consumption they observe in a single day is repeated in that same household on different days, or they might want to investigate the pattern across multiple households. Students will then use this to guide the formulation of a query to the dataset, which they will then visualise. Students will interpret across the set of visualisations they produce to answer a question. The focus in this trial will be on assessing the types of questions that students ask and whether they are good questions that are testable with the data, or not. A supplementary focus would be what are the technical or conceptual difficulties students encounter when conducting their own data analysis and how could this be better supported through technology? Students in this trial will be tasked with demonstrating additional data skills with the complex datasets, namely selecting, cleaning and visualising the data themselves. We will also aim to more formally validate some of our assumptions with regard to the use of narrative principles for framing queries and the use of PPDAC to support inquiry, specifically the idea of starting at the *conclusions* stage for existing data sets, then prompting students to conduct their own

data analysis to contextualise and consolidate the learning. To this end we will identify a set of formal criteria for assessing data literacy skills, such as ability to ask good questions from data, or ability to critique data, ability to write stories about data to contextualise it in terms of the real-world.

Finally, we will evaluate the approach across a number of classrooms and elicit feedback directly from participating students, which will inform the future design of the Urban Data School (UDS). We will expand the resources of the UDS to include activities focused on more general data analysis techniques, such as cleaning data, identifying outliers, critiquing data, combining data sets, or understanding issues of privacy. We will investigate the possibility to scale up the approach to teach data literacy in different contexts and, more crucially, to different parts of the UK.

7. CONCLUSIONS

An approach to teaching data skills in schools has been developed based on principles of narrative and inquiry-based learning. The data is presented in the context of a smart city topic, to support learners to make connections between the data and a real life environment and to provide an opportunity for learners to gain experience in using real, complex, large data sets. A set of materials has been developed for teaching data skills in the context of a topic about home energy consumption and generation from solar PV. These have been delivered as a concrete lesson plan in a primary school to 12 students aged 9-10 years. Students were observed as they undertook the tasks, to be able to make accurate interpretations of visualisations of real-life data sets and to tell good stories about what they showed. This suggests that taking an inquiry-led, narrative-based approach is suitable for teaching primary aged students some data skills, specifically those related to data interpretation. We propose to replicate this with older students who will undertake more complex data analysis and demonstrate additional data skills such as selecting, cleaning and visualising data. The trials are informing the development of a web-based platform called the Urban Data School which will provide teachers with access to big data sets and the lesson plans, tutorials and other resources required to use them in the classroom. We suggest that if, as we have demonstrated, young learners are capable of working with large data sets, this provides evidence for the argument to create a provision within the UK curriculum for this topic so that future generations will leave school with better levels of data literacy.

8. REFERENCES

- [1] Allmond, J. S., & Makar, K. 2010. Developing primary students' ability to pose questions in statistical investigations. In C. Reading (Ed.), *Proceedings of the Eighth International Conference on Teaching Statistics*, Ljubljana, Slovenia
- [2] Arnold, P. 2008. What about the P in the PPDAC cycle? An initial look at posing questions for statistical investigation. *Proceedings of the 11th International Congress of Mathematics Education*, Monterrey, Mexico
- [3] Bakhshi, H., Mateos-Garcia, J. and Whitby, A. 2014. Model Workers: How leading companies are recruiting and managing data talent. Retrieved May 8, 2015 from Nesta: <http://www.nesta.org.uk/publications/model-workers-how-leading-companies-are-recruiting-and-managing-data-talent>
- [4] Bourgeois, J., Linden, J., Kortuem, G., Price, B. and Rimmer, C. 2014. Conversations with my washing machine: an in-the-wild study of demand shifting with self-generated energy, *Ubicomp '14*, Seattle, US
- [5] Chatman, S. 1980. *Story and Discourse: Narrative Structure in Fiction and Film*. Cornell University Press, 1980.
- [6] Dewey, J. 1933. *How we think: a restatement of the relation of reflective thinking to the educative process*. Heath: Boston, MA
- [7] Haddadi, H, Howard, H., Chaudhry, A., Crowcroft, J., Madhavapeddy, A and Mortier, R. 2015. Personal Data: Thinking Inside the Box. Accepted paper at *Critical Alternatives 2015*, Aarhus, Denmark
- [8] House of Lords Select Committee on Digital Skills. 2015. Make or Break: The UK's Digital Future. Retrieved May 8, 2015 from: <http://www.publications.parliament.uk/pa/ld201415/ldselect/1ddigital/111/111.pdf> Accessed on 8th May 2014
- [9] Kastens, K. and Turrin, M. 2011. Geoscience data puzzles: developing students' ability to make meaning from data, *ED11C-04*, San Francisco, US
- [10] Krumhansl, R., Busey, A., Krumhansl, K., Foster, J. and Peach, C. 2013. Visualizing oceans of data: educational interface design. *Oceans*, San Diego, US
- [11] Maycotte, H. Data literacy - What it is and why none of us have it. 2014. Retrieved May 8, 2015 from Forbes: <http://www.forbes.com/sites/homaycotte/2014/10/28/data-literacy-what-it-is-and-why-none-of-us-have-it/>
- [12] McAuley, D., Rahemtulla, H., Goulding, J and Souch, C. 2014. How Open Data, data literacy and Linked Data will revolutionise higher education. Retrieved May 8, 2015 from Pearson blue skies: <http://pearsonblueskies.com/2011/how-open-data-data-literacy-and-linked-data-will-revolutionise-higher-education/>
- [13] Ojo, A., Curry, E. and Zeleti, F. 2015. A Tale of Open Data Innovations in Five Smart Cities. *International Conference on System Sciences*, Hawaii
- [14] Schank, R. 1990. *Tell me a story - a new look at real and artificial memory*, Macmillan, New York
- [15] Segel, E. and Heer, J. 2010. Narrative Visualization: Telling Stories with Data. 2010. InfoVis, Utah, US
- [16] Tech Partnership. *Employer insights: skills survey 2015*. Retrieved May 12, 2015 from: https://www.thetechpartnership.com/globalassets/pdfs/tec_employer_skill_survey_web.pdf
- [17] Tomashevsky, B. 1965. "Thematics." *Russian Formalist Criticism: Four Essays*. Comp. Lee T. Lemon and Marion J. Reis. Lincoln: University of Nebraska, 1965. 62--95. Print.
- [18] Vahey, P., Yarnall, L., Patton, C., Zalles, D. and Swan, K. 2006. Mathematizing Middle School: Results From a Cross-Disciplinary Study of Data Literacy. *American Educational Research Association Annual Conference*, San Francisco, CA
- [19] Wild, C., and Pfannkuch, M. 1999. Statistical thinking in empirical enquiry (with discussion). *International Statistical Review*, 67(3), 223-265
- [20] Wolff, A., Mulholland, P., Zdrahal, Z. and Joiner, R. 2007. Re-using digital narrative content in interactive games, *International Journal of Human-Computer Studies*, 65, 3, pp. 244