2012

Memeserver Ltd &
Political Innovation

Paul Evans

POLITICAL INNOVATION
Innovative Conversational Politics

TOWARDS A LOCAL DATA VISUALISATION DAY FOR SCHOOLS

An investigation into the value of bringing school pupils into the public debate around the visualization of open data, along with a summary of some of the practical considerations. This report includes original interviews with school teachers and a proposal based upon conversations with a range of governmental, non-governmental and private sector agencies.

Contents

Proposed initial circulation list:

- David Branch (Deloitte) dbranch@deloitte.co.uk, Emer Coleman (Government Digital Service, The Cabinet Office) emer.coleman@digital.cabinet-office.gov.uk, Andrew Collinge (London Data Store) andrew.collinge@london.gov.uk, Nick Jones (10 Downing Street) njones@no10.x.gsi.gov.uk, Emma Mulqueeny (Rewired State) emma@rewiredstate.org, Mark O'Neill (Government Digital Service, The Cabinet Office) mark.oneill@digital.cabinet-office.gov.uk, Chris Palmer (London Borough of Barnet) chris.palmer@barnet.gov.uk, Mark Pearson (NESTA) Mark.Pearson@nesta.org.uk

The intention of this report

This report has been compiled by Paul Evans of *Political Innovation* with a practical purpose; to provide a local authority with the information it needs to manage an experimental 'Open Data Visualisation Day' for school pupils.

This will include an outline of the some of the reasons that it would be worth taking this idea forward. It also looks at the practicalities of doing this along with a proposal to conduct a data visualisation day with a large number of pupils, working with experienced data analysts and visualisation specialists.

In preparing this report, a trial day was organised with school pupils from the London Borough of Barnet in which they spent a day working with data analysts from Deloitte LLP. The initial feedback from the schools concerned was very positive and it attracted some press attention[i]. As a follow-up, informal interviews were conducted with teachers from the schools concerned and some of the observations are included in this report.

Why is open data and visualisation important?

A wide range of work is being done around open data, and it is increasingly being seen as an important driver of renewal, both in government and in the wider economy[ii]. It is also seen as an area for entrepreneurial growth[iii].

There are opportunities to improve the quality of policymaking, resulting in more transparency and in less rigid and agency-led thinking[iv]. On all sides of the political spectrum, an opportunity to improve the efficiency of government at all levels has been identified[v].

There is also the opportunity to promote civic engagement – 'crowd-sourcing' public judgments on complex issues[vi].

There are also a range of concerns around participative innovation related to established models of Representative Democracy. It has been argued that – when public data was largely monopolized by civil servants – that it would be used in the public interest. By removing this monopoly, it could benefit pressure groups and organisations who have an interest in directing government spending[vii].

This report proposes to take those concerns very seriously. The integrity of elected politicians does not need to be compromised by well-managed participative exercises. In some models, participative decision-making could be portrayed as a rival to the decision-making conducted by elected representatives.

However, the opportunity here is to build an intelligent and useful conversation that increases public understanding of the issues and helps elected representatives – usually 'generalists' to improve their understanding of complex issues.

Democratic Renewal

Looking at it from the point of view of a local authority (particularly the communications team as well as the councillors) transparency and open data seem to have created a situation where the amount of time spent dealing with the *angriest* local residents goes up.

In principle, there is a political consensus that 'armchair auditors' are an adornment to democracy. They bring fresh eyes to complex decisions and could be a cost-free alternative to expensive management consultants.

On the other hand, *armchair auditors* are often viewed with suspicion by local politicians and the communications teams of local authorities. The term 'auditor' implies a neutrality and balance, but local authorities often see only non-neutral political activists who are selectively disrupting the aspects of the local authority's work that they don't like.

Of course, any democrat would answer that this is often a good thing. One would need a very strong case to discourage democratic dialogue, and policymaking that has more participants involved should be better than a more closed process. However, a democrat would also have concerns that this kind of relationship with active-citizenry can fundamentally undermine the duty of a democratically mandated organisation to address the concerns of the *whole* electorate – and not just the demographic groups that have the capacity to engage in this way.

It could be argued that Councils only have themselves to blame for this. If information has to be dragged out of them by Freedom of Information (FOI) requests, then only people with time on their hands will do it, and the information that is uncovered would even be selectively damaging to the political and administrative incumbents.

Similarly, if information isn't attractive to engage with, then only people who have the time to puzzle it through will do it. If information isn't made available in a format that allows other websites and forums to consume it, then the only people who will look at it are the *lean-forward activists* who are motivated to trawl the Council's website.

On the other hand, good, attractive, easy-to-read, well-presented information can – and will – be linked to from Facebook groups, Mumsnet, Netmums, hyperlocal sites and other relevant local forums. Often 'closed' or opaque information policies create hostile interest where it wouldn't exist ordinarily. This may be another example of 'The Streisand Effect'[viii] whereby information becomes more widely circulated if an attempt is made to suppress it.

This highlights our problem: How can we get a lot more of the key information that Local Authorities provide about themselves into an easy-to-understand, easy-to-engage-with, easy-to-share and easy-to-mashup format quickly and easily? How can this be done in a way that tells councillors things that they don't know as well as things that they may not want to know?

There is a possible answer to this – one that could have a game-changing potential.

School pupils need to improve their understanding of the manipulation and visualisation of data. They need to improve their presentation skills. They are less likely to conduct 'inductive' research as they are more likely to come from a representative sample (albeit with an obvious age-bias) of the local population. By encouraging schools – on a large scale – to request, tidy up, and visualise local data, we may have a way of tapping into what Clay Shirky describes[ix] as 'The Cognitive Surplus'

Open data and visualisation – how school pupils can help improve local democracy

This report will include some interviews with school teachers, outlining ways that open data and visualisation are at the heart of the school curriculum. Organising a day examining open data would undoubtedly provide those schools with a valuable resource to help promote learning in this important area.

However, there is an additional, intriguing reason for doing this.

Local authorities – and many other government bodies – are now expected to engage the public in policymaking to a greater degree than ever before. Yet, though this is increasingly a statutory obligation, many government officials freely admit that the thinking around effective participative policymaking is firmly in its infancy.

At the crudest level, policymakers find it hard to engage the public in complex questions and finding the optimal balance between original observation and wisdom from the general public, the contributions of experts and the demands of highly motivated interest groups can be very problematic.

In addition, all decisions are subject to the institutions of representative democracy and the permanent government institutions that statutorily manage public policymaking.

Local authorities need to engage with large numbers of people rather than the small proportion of the population who are the angriest – or those who want to push the council into doing something that is not necessarily in the interests of everybody living in the area.

The *'am I bothered?'* pyramid

After all, there will often be a small number of people who have a long-standing grievance with an organisation, or who feel very strongly about a subject that is under public consideration.

These are the people who send regular Freedom of Information requests. They are often also the people who are more likely to respond to consultations, write letters to local newspapers or attend public meetings.

Increasingly, they've also turned to blogging. They're at the top of the *'am I bothered?'* pyramid (above). This is a variation on Jakob Neilsen's observations on 'Participation Inequality'[x] though it should be noted that a recent BBC briefing has presented evidence that the 'commenter / lurker' ratio is shifting[xi] as web-users become accustomed to participation.

There are important democratic reasons to limit the influence of the 'red triangle' *active citizens*. Are they entitled to any more influence over public policy than those in the lower layers? Should their opinions or interests weigh any more heavily than everyone elses?

In some cases, government bodies almost have a duty to sidestep the *hard-to-avoids* to engage with the *hard-to-reach* people who also have the right to have their interests served.

But how can public bodies do this if they are spending a lot of time just answering the (legally enforceable) demands of this group?

One answer is to make it more interesting – to invest in information design, copywriting, play-based strategies of the kind outlined at a recent set of events on Political Innovation[xii].

Another would be to find a way of encouraging a large number of people who don't have a particular material or ethical interest in an issue to work together to come up with an agreed 'description of the problem' or a disinterested perspective on a particular issue.

A third opportunity is to find a way of engaging with active citizens in a way that is constructive – where their contribution can be negotiated with. There is an obvious problem with an unelected and energetic minority that is capable of compelling elected representatives to do something. But these same people may be very helpful to a local authority in helping to describe a complex problem.

From our exercise with the London Borough of Barnet and Deloitte, we looked at local transport and traffic issues as an example. In a policy area usually dominated by motorist / cyclist / environmentalist / business lobbies, a series of well-rehearsed arguments can be expected.

Our experience, however, was that pupils introduced a new dimension that no-one expected. Not because they were a new advocacy group (they weren't) but they noticed something without indulging in *inductive reasoning*.

One example was the way they examined the data around cycling / driving to school. Instead of looking at the data through the usual pro-cyclist / pro-motorist templates, the pupils picked up on the surprising result that school journeys as a car-passenger were more hazardous than they realised[xiii].

In another example, the pupils were able to evaluate the quality of journalism in evaluating a claim[xiv] that Barnet Borough has the most road deaths in London

There is one final reason for considering a form of policy deliberation that involves school pupils who are more disinterested in policy issues than the usual consultation respondents. It is an attribute that they can bring to policy-making that is valued by political scientists everywhere; the benefit of *uncertainty* to optimal policymaking. As Charles Darwin put it so succinctly;

 "Ignorance more frequently begets confidence than does knowledge."

School pupils, analysing local data can help to shield policymakers from the Dunning Kruger Effect[xv] - a problem observed by students of optimal policymaking where crude and simplistic proposals are advanced with an inappropriate level of confidence by organisations that have some disproportionate influence over policy decisions.

With millions of school-age pupils in the UK, a successful pilot around the analysis of local data could unlock what Clay Shirky describes as the cognitive surplus[xvi] – a large amount of human processing power that could be efficiently applied to big problems.

What does a schools open data day look like?

In March 2012, pupils from three schools in the London Borough of Barnet were invited to spend a day with the data analytics team from Deloitte LLP.

The day was broken down into a number of activities and the results can be seen in the report provided in Appendix A, 'One Day in March, 2012'.

1. Introducing the subject - showing pupils what is possible

For this exercise, attention-grabbing examples of visualisations that can change dramatically as data changes are shown to the pupils. For example, crime data was imported into a visualisation tool to show car-crime in different localities – including areas known to the pupils. It can be seen that – by selecting out bits of data and using it in a mapping tool - we can create and juxtapose visualisations to draw conclusions. We were able to show, for example, that there is a possible link between street-crime and street lighting.

The aim of this exercise is to...

- Show that data can be visualised, and that doing so enables us to extract meaning from what would be an unwieldy table of numbers
- Show that two pieces of data can be juxtaposed to test a hypothesis
- Show that the visuals can be manipulated by changing the numbers – giving a useful visual representation of what statistical changes can often mean in the real world
- Show that data can be illustrated in a traditional 'graph' form, but that it can also have geographical attributes that make it useful in conjuction with a map
- To show that this representation of data can provide a helpful conversation piece, enabling the group to discuss the question in hand more effectively

2. Understanding what is possible

We ask pupils what kind of problems they want to address and then lead a discussion on what kind of data they would need to do this. We give them a brief tour of the data that we know to be available including 'open data' sources on government-related websites. In addition, a few pieces of proprietary data that are procured in advance.

This also involves a discussion aimed at helping with critical thinking around what kind of data it is reasonable to expect to be available at no charge. It's good to ask pupils if it is likely that the data that they will need is collectable, and if it is likely to be collected and available to them at an affordable price. It's important for pupils to understand that there are some hypotheses that may not practically be tested using data for these reasons.

The aim of this exercise is to...

- Show that data is collected, and that – in some cases – it is available
- Help pupils understand what data they may need to start looking for in order to address a particular problem
- Understand what kind of data is available and what is likely to be collected – and thereby understand what hypotheses may be tested in this way (and which ones are less likely)

- To be able to think critically and understand that some data is available subject to some pricing or that exclusivity is applied for other reasons (confidentiality, privacy, legal reasons relating to data that may be reverse-engineered to compromise confidentiality, privacy etc)
- Familiarise pupils with the obvious sources of information
- Introduce 'lateral thinking' (i.e. "we may not be able to get the data we need, but there is other data that *could* help answer the question")

3. Data analysis– finding and formatting

At this point, pupils are split up into groups and are given a practical hands-on experience with direct help from analysts. This includes work around finding, cleaning up and formatting of data, perhaps getting the data into the kind of order that they want it in (using standard spreadsheet manipulation to sort data and export it to a CSV file). Year ten and eleven pupils were - in some cases - able to do this already.

The aim of this exercise is to...

- Introduce pupils to the basics of spreadsheet data-manipulation and the need to tidy up data
- Show some tidying up techniques and tools
- Identifying specific useful data-sets within a large pool of data

4. Data analysis - visualisation

With the help of the analysts, the data is processed using visualisation tools to enable the content to be visualised. Pupils are able to see the data being imported into the visualisation tools and can experiment with this and see what options are open to them.

The aim of this exercise is to...

- Show how data is added to different visualisation tools
- Show how changing parameters can change visualisation results

5. Drawing conclusions

After a period of experimentation, pupils are asked to draw up their conclusions. These are then written up, using screen-grabs and other 'exports' from the visualisation tools, and place into a presentation format (using Powerpoint or similar tools). These can be seen in Appendix A.

The aim of this exercise is to...

- do the main work of the day; to use the data and the tools to test a hypothesis,
- to identify useful conclusions, capture them and write up the results
- to 'grab' graphics and insert them into a presentation document or slide

The results of this exercise can be seen in Appendix A: 'One Day in March, 2012' (attached). It should be noted that this was the first exercise of its kind. It was conducted using a range of untried investigations that were chosen by the school pupils concerned.

6. Pupil presentations

At the end of the day, pupils present their findings and argue for the conclusions that they have drawn. In some cases, pupils have specific policy proposals that they wish to advocate. At the London Borough of Barnet / Deloitte day, pupil presentations were given to senior figures at Deloitte along with a (Cabinet member) Councillor from Barnet along with the Head of Communications, both of whom played the role of sceptical policymakers.

The aim of this exercise is to...

- Help with presentation and debating skills
- Understand the real-world application of this whole exercise – especially (in this case) in the context of local public policy

This was achieved in a day that started at 9.30am and finished at 3.30pm.

Data visualisations and the UK schools curriculum

There are a number of points at which an exercise using and visualising data could prove useful to schools in helping to develop their curriculum, to assist with existing elements of the curriculum, or as means of adding a new level of interest into the existing curriculum.

The obvious examples are in helping pupils to understand science methodology, and to give it a real-life context - helping pupils to grasp the scientific method.

Using data and visualisation creatively can also add a dimension to specific subject areas. In particular, schools are often asked to carry out 'controlled assessments' in which pupils conduct an investigation to address a particular problem. This is, specifically, an exercise in which pupils are expected to show an understanding of the methods used to analyse data in order to draw conclusions.

In many cases, many schools are doing the same controlled assessments at the same time. This could provide a very good opportunity for schools to focus on a particular area of study. Working together, schools within a local authority are could use an interesting practical experience around data-visualisation to improve their work around these controlled assessments.

In addition, teachers see benefits in helping them to build cross-curriculum learning. While this is often an approach that is valued, in practical terms, teachers and schools often find it hard to coordinate and maintain cross-curriculum learning and it may be the case that a school-wide data visualisation exercise could also help with this.

A number of teachers were interviewed - both individually and in groups. The interviews were informal and unstructured. The following observations about specific subject teaching were drawn.

Science

Some schools have a section of coursework called 'Practical Data Analysis' looking at data collected, understanding outliers and anomalous results and plotting these on graphs. Pupils are expected to plot results on 'suitable lines of best fit' - possibly curved, straight line or combination graphs that they then need to be able to interpret. They have to be able to reach conclusions.

They have to be able to explain how closely data matches a hypothesis and where it doesn't match. This is done in year 10 and year 11 as formal coursework, though pupils are looking at graphs from year seven onwards - it's one of the first things they look at when learning 'how science works' - one of the first units pupils do is 'planning and investigation, collecting data, plotting and analysing it.' - it becomes more complex and formal as they move through the school.

Geography & History

From year seven onwards, as with science, this is a key part of the curriculum. Pupils will have to test hypotheses, research and use statistics and other data to come to conclusions.

In year ten and eleven, pupils are doing formal pieces of 'controlled assessment' and they have to do their write-ups in the same way that they would in science classes.

In year seven, they are asked to do a 'local area study' - researching their local area, suggesting improvements on how they could change their local area. In year eight, they study population and migration, looking at - and understanding - statistics. In the past, some teachers had done similar work looking at local crime figures.

Also, studying tourism involves getting data on employment from the ONS website.

Teachers have a range of anecdotes showing specific uses of data. For example, the question of causation and correlation is illustrated by showing a graph of ice-cream sales juxtaposed with reported incidents of hay-fever, thereby helping pupils grasp the importance of questioning a 'causal link'.

Business Studies

Again, the controlled assessment of pupils involves them doing research around the viability of potential business ideas. Pupils use ONS data around deprivation, employment rates and different socio-economic groups. In 2011-12, schools have been asked to use this data to see if they should open (for example) a high-class Pizza restaurant, or alternatively, a cheaper pizza takeaway.

These assessments are often uniform ones and a number of different schools will be doing the same assessment concurrently. Open data projects could be very useful in supporting controlled assessments of this kind for year ten pupils.

Maths

In key stage three, teachers are keen to work in a more cross curriculum way, and they are keen to develop the kind of investigation skills that other subjects are developing.

There is a perceived need for this after some of this kind of work was recently taken out of the Key Stage Four curriculum, though some schools have an option to offer a GCSE in statistics.

Citizenship

Teachers see an obvious benefit to improving pupils familiarity with locality-based statistics for the study of citizenship, and in Personal Development Time, pupils are encouraged to improve their skills in presenting information and taking part in debates.

Business & Economics

Teachers often dig into the local authority website for data as local data is often particularly valued by teachers. In some cases, they are often frustrated by the lack of sourcing of some data that local authorities provide as pupils are also asked to cite and evaluate sources for credibility, accuracy or bias.

Sixth-form students will look at a wide range of data - often more national than local. Inflation data is used to help understand macro-economics. On more micro-economic questions, links between

- chocolate sales and obesity,
- petrol prices and congestion
- sugar prices and other commodities

...are looked at. One teacher outlined some observations that had been drawn about why differences in meat-pricing mean that McDonalds are more likely to advertise Chicken-based products than they would advertise burgers, for example.

> *"As science teachers we love our subjects, but few of us have been at the cutting edge of new research, so our data analysis skills are in need of sharpening at the very least. Therefore a chance to take part in real data analysis projects and for our students to gain such experience would be invaluable.*
>
> *The chance for cross-curricular work is always a good thing, but so hard to do when the curriculum is so demanding, and with the new focus on content and rigour from the Government it will become even more so."*
>
> **Jeremy Stone - Head of Science - The Compton School**

Organising a schools data visualisation day: practical questions

What kind of pupils are ideal for this?

A good mix of ages and abilities could be appropriate in some respects, but initially, it may be worth restricting this opportunity to pupils in years 10 – 13, and focussing it around the opportunity in the curriculum to support 'controlled assessments'.

When is the best time to do this?

Year ten and eleven teachers would be fairly unhappy about having pupils taken out of school at any time during the normal school week, given the demands of the curriculum, unless it is pre-planned and f it covers a part of the curriculum - for example, during the controlled assessment time.

What subject matter would be worth focusing upon?

Working with Deloitte, a number of subject areas (crime, traffic, accident, places of worship, etc) were explored. Drawing from this experience, it may make sense for the day to have a pre-determined focus in future. This would allow schools to prepare and it would give the day a sense of purpose.

The day could focus, perhaps, on two or three distinctive questions. For example, we could look in some detail at the data around traffic accidents.

We could introduce a new element and use it to investigate pupils perceptions around a particular subject area. For example, Voter Registration at 17 could be a useful one to examine.

Alternatively (or in addition) this idea could provide some synergy with local schools in their 'Controlled Assessments' – an important part of the school syllabus. They are fixed investigations that pupils are asked to conduct – testing a hypothesis, gathering and analysing evidence and drawing conclusions during a fixed observed period of time.

The actual exercises that schools are expected to conduct are disclosed to schools in September, so plans can be made after this point.

The appropriate level of support that a schools data visualisation day can offer to schools may be to provide help in one aspect of this (finding and preparing data perhaps) or by providing a condensed action-learning activity that prepares pupils for the tasks they will need to perform in a Controlled Assessment. It may be worth considering doing the data and visualisation day around a similar subject to one of the controlled assessments?

It may be the case that a single day may not be enough to provide all of the support needed for the actual Controlled Assessment in question.

On the other hand, an exercise like this could source better and more usable data than individual schools can, and it could be done in a more ambitious way. Perhaps the day could be about sourcing the data, or perhaps we could do something that is *similar* on the day rather than doing the same thing as the controlled assessment.

Pupils have to source the data during their own time and they often spend long periods on the ONS website. However, they don't have online access during their write-up period.

Pupils can take anything they want into the controlled assessment except for a pre-prepared outline of what they are going to write. The controlled assessment is about pulling the information together into a readable and useable format. Helping pupils to do the heavy lifting in advance of a controlled assessment may be a useful thing to do.

They would be less concerned about giving up some of the Year Nine pupils time though. Teachers were of the view that, if done during the school week, we'd get a cross section of the schools pupils.

The alternative is to consider a weekend activity. It will be more likely to attract the more able and enthusiastic pupils.

What resources would be needed?

We would need to identify a suitable building for this. This may be a school or college premises.

Ideally it will have ample Wi-Fi connectivity – and if possible, this will not be subject to some of the usual local government / educational site blocking systems. If these an be suspended temporarily, it would be useful.

We would need to ask pupils to bring their own laptops. We may need to ask them to install some visualisation software on the laptops in advance, but this may not be necessary.

We would need co-operation from a group of data analysts and open data specialists. For a one-off London-based exercise, this could be found from the staff some – or all – of the following:

- The London borough concerned
- The London Data Store
- Government Digital Services (Cabinet Office) and attendees and facilitators from Visual Camp[xvii]
- The team behind Rewired State[xviii]

- Deloitte LLP have expressed an interest in being involved in this project, subject to a wider agreement

All of these groups have been approached and have indicated an initial interest in discussing this further.

Report: Deloitte LLP data analysts meet pupils from the London Borough of Barnet

This is the transcript of an interview that the author of this report conducted with Adrian Tan – one of Deloitte's team.

Paul Evans: Adrian, we've agreed that this exercise is a starting point for involving school pupils in the London Borough of Barnet in the use of open data. We want to show them what is available, how it needs to be formatted so that it can be used and how it can be juxtaposed and visualised to draw some conclusions.

Personally, I'm interested in the democratic and participative aspects of this. I want to…

- Increase awareness among the pupils of the availability and potential uses for public data

- Consequently, increasing awareness more widely within the borough of this opportunity

- Illustrating ways that this method of working can tell us new things that we didn't know about the borough that we live in – thereby contributing to local policy development

- Highlighting problems that we encounter in working towards these aims (so that others can think about how these obstacles can be removed)

We'd like to look at what we can do with the public data that the government are pulling together from various sources onto a one stop shop at data.gov.uk along with anything else that we can get our hands on.

Adrian Tan: From a policymaking point of view, the potential use of public data is interesting.

- One could look at the data and identify clusters and patterns of occurrence. This is an evidence based approach which tells how events have actually taken place, rather than simply how we think they might have had.

- Armed with hindsight, we can then use the open data, with the right level of granularity, to mash it up with other sources to gain insights into the possible contributing factors. Open data enables valuable lessons to be learnt and unravels precious insights. What we need next is foresight

- Open data needs to be synergized with the right tools such as predictive analytics to proactively identify and address issues.

Fundamentally we are interested in how things will happen before they do, and not just simply why after they have.

PE: So what are we going to achieve by introducing school pupils to this idea?

AT: The reason it's worth involving school pupils is that – while the Open Data initiative is undoubtedly commendable,not everyone is capable of appreciating data.

A common drawback is the vast quantities of data one has to go through to identify patterns and insights. Manipulating and making sense of data requires experience and knowledge – it's not a skill

that can be acquired in a day. For a layperson, it's hard to know where to begin with raw data. It's not always straightforward even for data professionals.

We've been working on ways to ensure that the data isn't going to be overwhelming. Properly presented, it should engage people in an intellectually stimulating manner.

We will have an introductory day with the team at Deloitte, trying out some new and existing tools and we'll be writing this up once the day is complete.

PE: What can we do in one day with school pupils?

AT: We've been testing out new visualisation tools and we'd like to show some of these to the pupils. People who have the experience and statistical grasp can look at vast quantities of data quickly and become aware of the important details about the place that we live in, or alternatively we can put things that we broadly know under the microscope and see if we can learn something useful.

You need the right tools of course. Dull data is hard to engage with. But if it's presented in an intellectually stimulating way, it makes you think. We've always presented numbers as graphs, but today's software really allows you to get the numbers into a place where anyone – not just a statistician – can learn something interesting from them.

Visualisation provides a bird's eye view to an ocean of data and it also allows the user to put a magnifying glass on areas that matter most. Above all, visualisation allows people to lift themselves from the details, offer a fresh eye and possibly come up with novel ideas to address issues.

We want to show them what is available, how it needs to be formatted so that it can be used and how it can be juxtaposed and visualised to draw some conclusions/implications.

A good place to start with school pupils is to look at how they can use data to support what they're already learning in schools. Different groups of people will be interested in different kinds of data. If we can find the right questions, we can encourage school pupils to add to the sum of human knowledge – to explain their surroundings to the rest of their school. They could even explain their surroundings to their local authority and precipitate real change at a local level.

PE: So what useful data have we found so far?

AT: Well, let's start with traffic data. We can go here (insert link) to find out all sorts of information about local traffic accidents. We have the following information about each accident

1. Age Band of Casualty
2. Age Band of Driver
3. Car Passenger
4. Carriageway Hazards
5. Casualty Severity
6. Casualty Type
7. Control
8. Junction Detail
9. Junction Location

10. Light Conditions
11. Pedestrian Crossing – Human Control
12. Pedestrian Crossing – Physical Facilities
13. Pedestrian Location
14. Road Surface Conditions
15. Road Type
16. Sex of Casualty
17. Sex of Driver
18. Special Conditions at Site
19. Speed Limit
20. Vehicle Manoeuvre
21. Weather Conditions

There is more information than this, but I had to do some cleaning up and I had to exclude some information that was simply not in a useable format. The 21 streams listed here are the ones I could use usefully.

So are there combinations of 'road type' and 'speed limit' that are particularly likely to result in accidents or the severity of accidents? How far does weather conditions affect the dangerousness of a particular type of road?

PE: So we can plot these on a map and offer predictions about where accidents are likely to happen? Can we say 'when it's rainy and dark, you should particularly avoid this stretch of road'? Will school pupils be able to print off maps that they can show to their school, providing useful pedestrian warnings?

AT: As the data came with longitude and latitude figures, we were able to plot these on a map.

As for predicting about where accidents are more likely to happen, we have to look at this with a statistician's eye. If most of the accidents happen in fine weather, we can't necessarily conclude that bad weather has nothing to do with accidents for all of the obvious reasons – especially if there are no bad-weather days in the period we're sampling. And so on. We have to weight figures accordingly and not succumb to any of the usual cognitive biases.

This still leaves us with a lot to play with though. The interesting point about this set of data is the level of granularity. For each accident, there is record of whether it is slight, serious or fatal. This piece of information is crucial to a predictive approach. We could look at, within each category of vehicles type, the percentage of accidents that have resulted in a fatality and compare it across other vehicles. So we weren't surprised when we found that accidents involving motorcycles of over 500 cc were twice as likely to result in a fatality than other vehicle types.

Although this finding is intuitive and nothing new, by using this 'granular insights' approach, we could apply this methodology to other data fields and predict fatalities as well as accidents. Obviously, this is useful information for any policy-maker. Using this methodology, we hope school pupils can print off maps that they can show to their school on a number of variables, providing useful driver and pedestrian warnings.

PE: What else is interesting about the Accidents data?

AT: If we can draw data from several sources, we can provide a more useful picture. We can find things we hadn't even thought of looking for. However, this isn't always that easy. Matching up data may not always work. Different sources can granularise their data in different ways so we aren't always comparing apples with apples. Different time frames (data collected over different periods of ime) means we are unable to draw conclusive inferences about what is actually driving the behavior /outcome hat we are interested in. The Accidents data is an exception. Weather, road condition, junction type, accident severity etc is all recorded at the same time. But once we juxapose it with other information, it becomes trickier.

PE: OK. I know we'll be looking at other bits of information around crime and health, but let's stick with the Accidents data for now. Is there anything we could d to take it further than just a day event?

AT: We think that uploading the data used on that day onto a public server where the pupils can continue to play with it and share it with their family and friends will be useful. Having data on a public server means people can share a view. Sharing a visualisation helps to create a shared view of a situation and draws members of the community together and aligns thinking on how we can improve on things that matters most to us.

PE: And can councils and councillors get anything from this?

AT: In an ideal world, the pupils would get this data together and come out with recommendations that could be taken to a council meeting. In reality, that probably won't happen yet. What we're doing here, for now, is stimulating thinking. It should be of interest to councillors that school pupils can draw interesting conclusions from data. It shows councils how they can raise their game. It also illustrates the need to collect, store and share better data.

It will also result in an instructive process where we all understand statistics a little better. The dialogue between people crunching data and people who have to make decisions on it will always be interesting. We all have to learn various golden rules about how correlation doesn't guarantee a cause, or how common biasses skew findings and mislead us all.

But we need to make a start – hopefully we'll know a lot more once we've spent a day mashing data around with a group of school pupils.

Acknowledgements:

References

i http://www.standard.co.uk/news/education/pupils-taught-to-tackle-information-overload-7585152.html

ii http://data.gov.uk/opendataconsultation/the-six-opportunities-of-open-data

iii http://bdaily.co.uk/news/technology/22-05-2012/open-data-institute-to-exploit-data-business-opportunities/

iv http://www.guardian.co.uk/public-leaders-network/blog/2012/jul/06/open-data-white-paper

v http://www.computerweekly.com/news/2240159056/Whitehall-could-save-33bn-with-big-data-analytics-says-Policy-Exchange

vi http://www.politicalinnovation.org/2012/01/crowdsourcing-analysis-for-policymakers-what-could-possibly-go-wrong/

vii http://whimsley.typepad.com/whimsley/2012/05/open-data-movement-redux-tribes-and-contradictions.html#sec-5

viii http://en.wikipedia.org/wiki/Streisand_effect

ix http://www.ted.com/talks/clay_shirky_how_cognitive_surplus_will_change_the_world.html

x http://www.useit.com/alertbox/participation_inequality.html

xi http://www.bbc.co.uk/blogs/bbcinternet/2012/05/bbc_online_briefing_spring_201_1.html

xii http://politicalinnovation.eventbrite.com/

xiii See the findings of Elizabeth Akamo and Eve Harding in the 'One Day in March, 2012' document (Appendix A)

xiv http://www.times-series.co.uk/news/9029136.Council__high_road_injury_tally__inevitable_/

xv http://en.wikipedia.org/wiki/Dunning%E2%80%93Kruger_effect

xvi http://www.ted.com/talks/clay_shirky_how_cognitive_surplus_will_change_the_world.html

xvii http://oursociety.org.uk/group/visualcamp

xviii http://rewiredstate.org/