Interview: Alon Peled on the Public Sector Information Exchange, avoiding disasters, and big data

By Democratic Audit UK   2016-6-15

Alon Peled is the man behind the Public Sector Information Exchange, a new digital tool which will allow citizens to access state agency data and compare it across different countries and continents. Last year, *Sean Kippin* spoke to Alon about both this project specifically and the potential for big data more generally.

Part 1 of this interview can be found here.
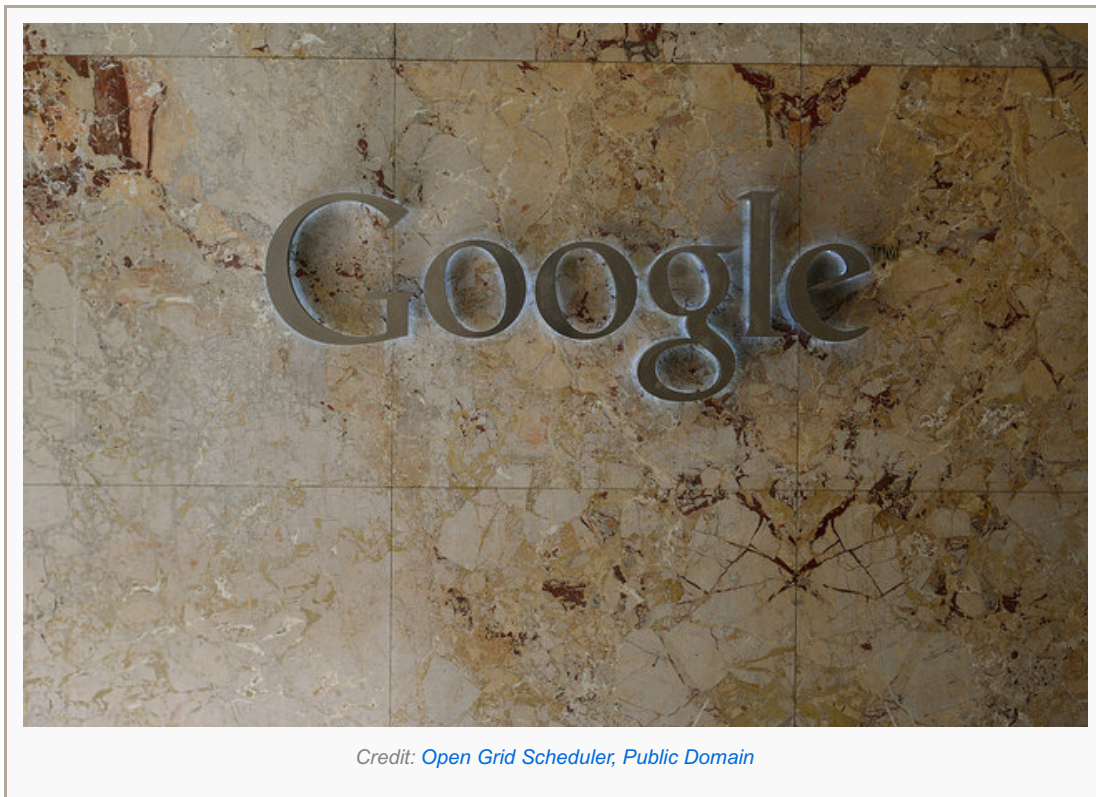
Credit: Open Grid Scheduler, Public Domain

In your presentation, or the video that preceded it, it showed dramatic scenes depicting 9/11 and Hurricaine Katrina. I was wondering how, in a fully implemented PSIE, if crises like this could be averted by better data-sharing. How would that work?

Yes, for example, in the first chapter of my book, I discuss the major disasters that have captured the public's imagination beginning even before 2000 and argue that they all have some origin in public information sharing failure. On 9/11 we had planes crashing into buildings even though perfect information on the lead hijacker was in the possession of the US intelligence services. In Japan, fishermen were directed into a radioactive situation because their national agencies weren't sharing the relevant information in a timely manner. So, you can see the importance of data sharing across agencies and its potential to improve the way that agencies deal with countless different situations.

The other thing that is important is that there is an assumption that data has a lot of natural security problems and that there are worrying implications for privacy. The truth of the matter is that, when you dive into the public data of many state agencies in many countries across all of the world's continents, one of the things you quickly come to realise is that the majority of government data is not about national security or personal privacy, its generally data

about the environment, disaster relief and economic development, for example. So what I'm arguing is that before we get to the 1% of problematic data, can we focus on the vast majority of it that doesn't have any security or privacy implications and could be used to create better outcomes. With that non-harmful data, I want to move towards a situation where the initial assumption is that if agencies don't share it, then they're doing something wrong. A situation where there are far fewer reasons not to share.

When it comes to that sensitive data, high silo walls might make more sense, particularly in a national security context. But with Hurricane Katrina – bodies floated in the water for three days, because agencies would not share the relevant GPS information about the location of the bodies with one another! Does that sound reasonable to you? There are no excuses here. What needs to happen is that we need to put in place is a mechanism to share this data, and you will start seeing, once that happens, the quality and the availability of the data improve, with the added benefit that maybe we'll see fewer disasters, or that when disaster happens they will be mitigated much faster.

What's next for the PSIE? Its active and its functioning, but is there any possibility that governments are going to begin actually using it?

I suppose we have an 'embarrassment of riches' as you'd say in English. I took a detour from the main stream of my research – which is to say that I would very much like to come and finish the prototype and go and work with agencies and try and install it, and I have some interesting potential opportunities to do just that. But, this detour is basically the result of conversations that that I had with people who said 'when you are building your test data, you have the potential to actually solve an equally important problem, the problem of how to make metadata accessible to systems'. So I try to think about it in terms of a triangle; the three corners of the triangle are a) government; b) corporations; and c) civil society. If I ask 'who is playing the big data game' then we can see that corporations are playing that game. Government are playing that game. But who is not playing? Civil society! The nature of the PSIE project is about putting big data in the hands of civil society. I am one of the guys saying 'I will stop for a second dealing with government information' in order to arm citizens with big data so they can use it.

I'll give you one example that I hope you'll find interesting. I have 4 gigabytes of metadata, some I developed and some I collected. That has its origin in 10 *terabytes* of government information. This process is basically taking very big data, and turning into small, usable, and useful data for the public to use. Lots of academics and researchers approach me and say 'we want to play with your search engine data visualisation' but that's an academic project, and the world needs useful data. So essentially I am taking a small break in the big PSIE project and I'm trying to make that data accessible to citizens and to society. When I do that, I am assuming that the traffic will come in, because it'll the only place on earth where you can find real comparative government data. If you want to compare say, Paris to London, you don't have much of a choice. You can go to either the Paris open data portal or the London website, and collect the data, and then you have to go to the European Union website and manually put it all together. However, because mine is integrated and has the data from all of these places and more, in all languages I expect a lot of traffic from people who want to compare these different things. So I will be in a position to, for the first time in history, capture the different uses people have for data they are searching for. Then I will publish that, too! That meets the demand and supply sides both met. So that is my current pet project.

Presumably with citizens having this incredible range of data, there are big implications for political choices that people make, and the eventual shape that society takes. Do you think this could be genuinely transformative in the way that civil society interacts with government?

It could be. If you put the right tools in the hands of citizens. Agencies don't release the most valuable data – if they have a way to sell it, they'll sell it. The reason eBay, for example, does so well, is because of the 'Long Tail'. They've found a way for people all around the world to sell what they have in their attic. Multiply that by 8 billion people, and eBay created an extremely profitable situation. Amazon did the same thing with books, and Google did the same thing with keywords and with spelling mistakes and with voice recognition.

There is enormous value in aggregating all of this data. I plan to do the same with open government data. In London

you don't necessarily need this service to have access to high quality data – but what about, say, fishing village in remote areas? What about the tens of thousands of cities who only release a tiny percentage of their data? If you had a way to aggregate the 'Long Tail' in the same way that Amazon does on books or eBay on the contents of peoples attics, then all of a sudden, even if governments don't release the most important information for researchers and entrepreneurs in a comparative way, they'll be a way to create something better. The essence of my software is aggregating that 'Long Tail'.

—

This is part two of a two-part interview. This post represents the views of the interviewee and not those of Democratic Audit or the LSE. Please read our comments policy before posting.

—-

Alon Peled is associate professor and political scientist at the Hebrew University of Jerusalem, Israel. He is the author ofTraversing Digital Babel, and his website can be found here.

This post was based on a presentation given by the author at the 2015 CeDem conference in Krems, Austria