# Making data, making sense

by Usman Haque


First published in ja+u (Japan Architecture + Urbanism), November 2014 – Data-Driven Cities – https://www.japlusu.com/shop/product/au-201411

'Smart City' and 'Big Data' are two phrases we hear about a lot these days, often in the same context: how technology can support urban processes. Both have a goal of using data to optimise cities, making them more convenient, predictable, efficient and secure. However, by seeing cities as problems waiting to be solved, these two techno-centric perspectives often ignore what it is that actually makes a city valuable and meaningful: its hyper-local serendipities, idiosyncrasies, unpredictabilities and poetries. If you consider a city as an engine for the exchange of unexpected ideas that manages to integrate across millions of perspectives, opinions, disagreements, goals and aspirations then the suggestion that an all-encompassing objective data infrastructure can help optimise it seems suspicious.

Data is not just 'out there' waiting to be captured like butterflies. It is crafted, collated and curated by someone or something (with someone behind it) that had a reason for doing so – a hypothesis and methodology for testing that hypothesis that informed decisions about measuring 'this' not 'that' and 'here' not 'there'. In order to make sense of data, 'Big Data' suggests that quantity trumps quality, that authorship biases will cancel out at scale, and that we will somehow know more if we can only measure more – all of which go against much of the last hundred years of the philosophy of science and society which demonstrates the paradoxes that arise when you try to remove observers from the observation. Making sense of data requires understanding its context: just as it is with making sense of people, so it is with making sense of data. We need to be able to push data, converse with it, prod it, interview it and build upon it in order to make any sense of it.

Data is meaningful if we have some way to act upon it. Otherwise, we are mere spectators. This is one of the most problematic aspects of the current Big Data fetish, which appears to treat data as an unquestionable justification for itself, rather than as a proxy for things that we actually want to understand or probe.

You generally can't put yourself into a dataset visualisation, tell it a little about yourself, and nudge it towards a better understanding of the questions you want to ask of it. If we are satisfied with mere data, datasets or visualizations as the end goal – rather than all the contextual complexity behind by whom, why and how it was collected, what was excluded, and how we can build upon it – then we are oblivious to all the multivalent inputs inherent within one-dimensional datasets.

Data doesn't need to be numeric, digital or electronic; it's anything that helps you make an assessment. In many senses, if it is non-digital then it necessarily integrates a whole host of other phenomena, providing a much deeper, if more complex, proxy.

A wonderful example of this was an air quality experiment led by Professor Barbara Maher, of Lancaster University. Thirty potted birch trees were placed directly outside the doors of four houses. Four other households, acting as control subjects, had no trees placed outside.

A major innovation in the experiment was that levels of particulate pollution were evaluated by collecting dust particles that settled on television screens, which had been wiped clean at the beginning of the experiment, and comparing the two sets of households to see which had amassed more particulate. Without resorting to a digital sensor sitting on a mantelpiece, the scientists and inhabitants alike could see and feel firsthand the difference trees made to the level of particulate.

One of the best ways to make data more meaningful is to make it yourself. Measure something – your body, your home, your neighbourhood – and it helps you not only understand something about it, but more importantly, it helps you figure out the questions you want to ask and the hypotheses you want to assess. Measuring something yourself (the way your body temperature fluctuates; the cycles of noise in your neighbourhood; or even dust on a television screen) means you can better decide why and what you might do to affect or act upon it.

A city hackathon brings dozens, if not hundreds, of software developers together for a short space of time to work for free on government-approved historical datasets. This is all well and good, but is it actually so transformative to work on something without questioning *why* and *how was the data collected? Which data was excluded? How would affecting its inputs alter real-time feedback?*

When you join others to measure something, you make meaning by having conversations about the data you are collecting. Sensemaking becomes a collective activity. You don't even need to be using the same measuring equipment, you just need to be able to talk about what you're doing with each other. "I'm measuring air quality," you say. "Well, I'm recording atmospheric humidity levels," says your neighbour. Have a discussion and you'll start to build up an intuition of how they correlate. Or even better, look at ways of affecting them together, ideally for the better.

The most important aspect of making data more meaningful is to experience it, somehow, in situ. Even if you were not part of the process of collecting a dataset, by being near to where and when it was captured, you are far more likely to integrate all the unspoken, ambient, implicit, informal and unrecorded 'metadata' that datasets and visualisations strip out for the sake of quantitative objectivity.

To stand in a space, a neighbourhood or a city and experience its swirling stimuli, while simultaneously interrogating, prodding and affecting a dataset, provides you with the kind of multivalence that is absolutely crucial to constructing any useful meaning. And you are far more likely to be held accountable, and to hold others accountable, for making use of the data in any decision making process.

There is a reason of course that capturing top-down data is so seductive to those who have to make decisions about cities. It enables them to outsource responsibility to automated processes, ones that enable them to say, "It wasn't me, it was the data!" Smart City IT vendors play into this because their business models have focused on pitching OPEX savings through automated processes. However, in an age of networked urban environments and ubiquitous connected devices, we are all generating data that will be handled and manipulated by processes that belong to other people and companies. It's vitally important for us, citizens, to develop a more sensitive relationship to our data, particularly given the growing tendency to abdicate decision making to algorithms buried inside software.

We launched Thingful (thingful.net), a search engine and discussion platform for the Internet of Things, in order for data sharing to be truly transformative. People need tools to find, discuss, make sense of and share the data that their connected devices and networked environments are generating. Data is now regarded as a valuable asset. But it is only valuable if you know where to find it; who created it; how trustable it is; how it compares to similar data and why it was captured in the first place. This is why we created the Thingful sharing platform.

Private data, on the other hand, is more sensitive and often personal in nature, so we also provide tools that allow owners control over access to their data. Some may choose only to share data with trusted third parties, or filter out location and other sensitive information. Others won't want to be found at all. For the Internet of Things to positively transform our cities, people need the tools to retain control over their data and to be able to discuss, make sense of and share it with those they trust.

By putting people back at the centre of analyzing data, we wish to stress that data is not ‘out there’ waiting to be discovered, captured and exploited. Data, meaning, decisions – they’re for us to craft, create and cross-examine for ourselves.



First published in ja+u (Japan Architecture + Urbanism), November 2014 – Data-Driven Cities – https://www.japlusu.com/shop/product/au-201411