

A byte too far?

Dave Marples comes over all nostalgic for a time when collecting data didn't have so many potentially harmful repercussions

Back when I was a boy, anyone could go into a shop, buy what they wanted and wander out, safe in the knowledge that the transaction was private between you and the shopkeeper, with only the receipt as a lasting record of the transaction. At the same time CCTV was something you pretty much only saw in episodes of futuristic TV programmes (such as Dr. Who) and facial recognition, speech-to-text and any other number of real-world data harvesting techniques were firmly constrained to the realm of science fiction. Storage limitations made the squirreling of huge quantities of data for later analysis prohibitively expensive (my first hard drive had a 32MB memory) and we were selective about the data we actually retained, choosing it carefully, based on information content and the things we'd be more likely to need later.

Before you conclude that Dave is off on one of his annual rants, allow me to explain what this has to do with Intelligent Transport Systems.

Well, let's fast forward to the present day

and we've reached a world where storage is so cheap that it can, for most purposes, be considered free – the smartphone in my pocket contains 500 times more storage than that first hard drive! The engineers' default action, particularly during systems development, is to record data "in case it's useful" with precious little regard to whether it is actually essential to the task at hand or not. The result is that everywhere we turn data is being squirreled away for later; in the UK we have somewhere between 1.85m and 4.2m CCTV cameras (a large discrepancy, admittedly, but it depends on who you believe and which newspapers you read), a fair proportion of which are recorded and, until recently, there was good chance the smartphone in your pocket was recording historic location information. There are caches and records made of many of your computer communications and phone calls, whether you comfortable with that or not.

That's not to say that logging all this data is not useful – how many crimes have been solved using CCTV evidence?

How well would ISP networks perform if they had no historical data to base their provisioning decisions on? Unfortunately the sheer volume of data being recorded means there's the potential for us to become the most highly monitored society that's ever existed by the gradual acceptance of data logging and the convenient benefits that historic records bring with them.

MAINTAINING THE ACT

In the UK at least we've got a pretty reasonable set of legislation to address these issues; the 1998 Data Protection Act (DPA) which enacts the 1995 European Directive on Data Protection. This act constrains the data that may be collected and retained in relation to individuals to ensure that it is necessary, is only used for the purpose for which it was collected, is kept secure and does not leak to jurisdictions that are not subject to the same constraints.

Nevertheless, people are starting to become nervous about this trend towards increased data storage – and it's not just



“Quantitative changes in data availability lead to qualitative changes in privacy. Most people haven’t realised that yet”

the kind of people who always worry about the sky falling in either. I was chatting to a doctor friend of mine the other day about his concerns about the location and access of medical records. He has explicitly forbidden his own doctor from uploading his medical records to off-site storage operated by the National Health Service because he isn’t comfortable with the safeguards and long-term protection of the data that’s there, or to put it another way he simply doesn’t trust what will happen next year, or the year after, or perhaps the one after that. You can imagine the fun that ensued when his doctor’s practice moved to a hosted patient record system. So, given that ITS is all about the remote sensing and processing of data, what does all of this mean?

Well, one thing is for sure that is really only just starting to be understood. Quantitative changes in data availability lead to qualitative changes in privacy. Most people haven’t realised that yet.

Let me give you an example: If I were to stand at the end of your road and post all of your movements, such as every time you left the house, onto the Internet, would you be concerned? The parallel with ITS is obvious: people say “Of course we can read your number plate, it’s on the front of your car and you’re in public, so why shouldn’t we?” Well, that’s fine, but storing that data permanently without due consideration for the things that can be inferred from it is little short of reckless and certainly within the interest of the DPA.

And don’t think information mining constraints imposed by current technical capabilities will protect you either. A few months back I got into discussion with one of my son’s clubs about the posting of photos on Facebook. The poster’s argument was that it wasn’t an issue as there were no individual names associated with the pictures... that was fine at the

time, but a few months later Facebook introduced facial recognition, and the game immediately changes. That same data can now be processed in a whole variety of new ways, inferring new things that were never previously intended or expected to be revealed.

BARE NECESSITIES

ITS systems potentially need to collect a lot of data just in order to make them work. I don’t for one second advocate that we shouldn’t collect data, nor should we be afraid to retain it when appropriate, but we do need to be careful to go back to first principles, to consider if we really do need to keep that particular record or this particular tag... to have that 32 MB hard drive virtually in our heads during design time. In modern Information Technology there is a distinct trend to grab as much data as you can lay your hands on just in case it’s useful later. Sooner or later people will come to realise the consequences of that data hoarding – and ITS needs to be as far away from that storm when it hits as possible.

I’ve lost count of the number of times I’ve been asked to develop a system with technical constraints on it to meet some data protection constraint – hashing car number plates to prevent extraction of the original number, or perhaps anonymising use records to

prevent identification of individuals.


This is good engineering and I’m very happy to comply (despite my engineering gene wanting to log everything) and we have all heard stories of data being lost on the bus, or deliberately accessed by errant journalists. When those things happen I’d like know that nothing of value can be inferred from what’s been leaked.

The ITS world would be more or less non-existent without its data, but we need to be very careful indeed about exactly what we log and constantly work to minimise the data we need to retain – and that’s not just to meet the requirements of the Data Protection Act or some other national equivalent, but because we have evolved over millions of years to be safe and secure in the knowledge that there’s a space at the back of the cave where we can be completely private and unmonitored.

I hate to think what will happen to our society when that is no longer true. 🙄



fyi

 Dr Dave Marples is Chief Scientist for Technolution, BV

 dave.marples@technolution.eu

 www.technolution.eu

 For previous articles by this author, enter MARPLES into the search box at thinkinghighways.com