14

# Traces and Shadows

Growing up in the 1970s and 1980s, I lived in a fairly analogue and small data world. All of our household appliances were electro-mechanical. My homework was paper-based and marked by hand. If I wanted to discover information, I went to the library and searched through hard copy books. Our car was purely mechanical, with no digital components or network links. I listened to music via the radio or by playing vinyl records or tape cassettes, and television consisted of three then four channels. Communication was by written letter and a landline phone. Undoubtedly, I appeared in a few key government databases, but most of my education, health and welfare records were stored in paper files.

There were some hints of the digital world to come. In the late 1970s, my parents bought a clone games machine that enabled the video game *Pong* to be played on the television,[1] and in 1981 I received a ZX81 personal computer as a Christmas present. It had 1K of memory (16K with a booster block).[2] To play games I first had to type in the programs then save them onto a tape cassette. A couple of years later, my parents bought a Spectrum computer for the family, which had slightly more local memory (16K, expandable to 128K), colour graphics, and you could buy pre-made games on cassette.[3] In the mid-1980s, my father had a satellite phone installed in his company car so he could be contacted when he was driving around the country to visit work sites.[4] I left home for university in 1988. The library catalogue was still mostly card based, but it was possible to do some electronic search for items. My essays were handwritten or typed, and communication with staff and departments was via

letters and noticeboards. In 1989, I first accessed the internet, still in the pre-web era, and had my first email account in the same year, though I barely used it as few other people I knew had an address.

In the 1990s, everything seemed to change. When I started my Master's degree in GIS in 1991, my parents bought me my first personal computer. For £900 I got a 286 IBM clone with 1Mb of internal memory and 16Mb hard drive. It was painfully slow to use, and if I set it to run a program that would rotate a small, basic array (for example a portion of a remote sensing image) 90 degrees it would take hours (an operation that is now complete before I've fully removed my finger from the mouse button). My lecturers, nearly all of whom were computer-savvy, communicated with the class using email, and I would use the chat facility to real-time talk with other students on the campus network (but not off-site). In 1992, I had my first encounter with a computer virus, which corrupted my thesis files (though thankfully I had week-old versions backed up). My digital data footprint at this time was still tiny. I had a bank card which I used in ATMs, I barely explored the internet, and most communication and record-keeping was still paper based. I was certainly identified by a suite of indexical codes[5] (for example, national insurance number, bank account, credit card, passport, loyalty cards, permanent address, postcode), but the digital data shadow attached to these was relatively circumscribed and added to sporadically.

The World Wide Web (WWW) started to change things by making information accessible across the internet through an easy-to-use, intuitive graphical interface. I first accessed the WWW using the Mosaic browser in 1993/94. In the same period, I invested in a home internet connection through a modem attached to the phone line. In 1995, I published my first article about the internet that detailed the geographical resources and data becoming available online.[6] I also made my first online purchase. Using the internet, I started to leave digital traces in my wake – locations, interests and movements (websites, hypertext links), interactions (email, chat) and transactions (purchases). In everyday life, my digital data shadow was also growing through the use of debit, credit and store loyalty cards, and captured in

government databases which were increasingly digital. By the late 1990s, mobile phones using 2G cellular networks were starting to become commonplace, though I didn't own one until the mid-2000s. I started to become fascinated with the transformation occurring, and my first book, *Cyberspace*, published in 1998, discussed the ways in which the culture, politics, administration and economy of everyday life were rapidly becoming mediated by internet technologies.

The speed at which digital technologies became embedded into the fabric of our homes, work and public spaces, and started to thoroughly augment and mediate everyday life – labour, travel, communication, consumption, play and leisure – is quite startling. In 2005, Martin Dodge and myself had two articles published that charted the extent of the changes. The first plotted day-in-the-life vignettes of three people's everyday encounters with digital technology from when they woke up to when they went to sleep at night.[7] Despite having very different backgrounds, lifestyles and jobs, each person encountered digitally mediated objects, systems, processes, interactions and transactions throughout their day and in different environments. In fact, we concluded that it was all but impossible to live a digitally free life and not leave data footprints[8] and cast data shadows.[9]

Running tandem to the creation of digital lifestyles was the datafication[10] of everyday life. In the space of 15 years, I'd gone from creating very narrow, selective data traces, to producing streams of varied and rich data. This was evident in the second paper which examined the various ways in which digital data was being generated and tracked using indexical codes about people, but also objects (using barcodes, RFID chips, manufacturer serial numbers, vehicle licence plates and so on), transactions (order numbers, shipping numbers), interactions (email ID, message ID) and territories (postcodes, grid references, latitude and longitude coordinates), and how these data were being used to govern people and manage organizations. To illustrate our argument we examined the data traces held within a typical wallet (which was actually Martin's wallet on 14 July 2003).[11] In the wallet was a variety of plastic cards, most with magnetic strips and/or RFID chips that stored key data, and pieces of paper. The cards included three library cards, a British rail card, a bank ATM/debit card,

a credit card, staff card, door access card and a donor card, each with its own unique ID code, and in a couple of cases a photo. The printed paper were mostly receipts, each of which included a merchant code, authorization code, till receipt number, and a portion of the credit card number used for purchase, some with printed barcodes on them. There was also a business card (that detailed work address, phone number, email and web address), a bus ticket and train ticket (transaction ID), Post-it notes with pin and telephone numbers written on them, and cash (each note having unique ID number). Each of us and our activities were being captured in data traces.

As the 2000s progressed, this datafication intensified through the continued application of networked digital technologies to every aspect of life. Systems and processes that were reliant on mechanical or electro-mechanical systems were augmented or replaced with digital counterparts.[12] Cars, for example, started to be fitted with sensors and digital controllers, using real-time data to mediate the driving experience.[13] Objects and systems became 'smart', using data and algorithms to anticipate and be reactive to use. They also started to log their own use, and if networked, share that log with developers, manufacturers and service providers.[14] The transition to Web 2.0, the introduction of the smartphone, online purchasing, and e-governance, in particular, deepened personal datafication.

Up until around 2003, the WWW was largely a broadcast media. Companies and those technically savvy enough created websites through which internet users could access a vast collection of information and purchase goods and services. The emphasis was on consumption rather than participation. In the early 2000s, however, a transition to what was termed Web 2.0 took place in which the production of web content was diversified.[15] A new suite of media tools enabled internet users to easily author and share their own content and to interact with one another. These included social networking sites such as Facebook, blogs, synchronous microblogs such as Twitter, shared gaming spaces and multiuser virtual worlds, mash-ups, media and code distribution services such as YouTube, Flickr, and SourceForge, peer-to-peer file sharing via protocols like BitTorrent, and social tagging and bookmarking. The internet

became a read+write media, in which people added value to sites as they used them.[16] In the process, users of these services shared their thoughts, opinions and values, as well as revealing their social networks.

Smartphones started to appear in the early 2000s, with the launch of the iPhone in 2007 creating a tipping point into mass use. Merging the functions of a personal computer with mobile phone, a smartphone provides powerful, networked computation. On the move, a user has access to the internet, email, a camera and a diverse range of apps. Along with the data gathered by the phone itself through its range of sensors, the data transmitted through its use mean that a smartphone constitutes a spy in our pockets, generating granular and detailed sets of data about us. Purchases and other transactions, such as online banking, are routinely conducted using smartphones, as well other personal computers, providing detailed traces of consumption patterns and lifestyle. The shift to using digital systems and e-governance at all levels of government mean that we often interface with public agencies through the internet, sharing required information by filling out forms. Key personal data are piped directly into state databases, with algorithms working on these data to ensure we are compliant with legal responsibilities and receive appropriate benefits and services. Other personal data are generated by mass surveillance via cameras, phone and Wi-Fi tracking, travel cards, and networks of transponders and sensors.

We now live in a world of continuous data production, since smart systems generate data in real time. Data streams off your smartphone, smart TV, smart car, laptop and tablet, and any other networked computational device, even when you are not directly using them. As well as seeking permission to use the camera, microphone and other sensors, Android apps can look to access a range of data stored on the phone[17] including device information, email, phone and message logs, photos and sound recordings, internet traffic and clickstreams, Wi-Fi connections, app activity and data usage, battery details,[18] GPS logs[19] and telephony information. Data are being harvested by the device manufacturer, operating system provider and app companies.

If you want to get a sense of how much data is being captured about you it is instructive to look at the data holdings of Google

and Facebook. Both companies provide options to let you look at and delete all the data they hold about you.[20] Through their various services and subsidiary companies (for example, Facebook also own WhatsApp and Instagram, plus many others), these two companies generate an enormous volume of data about their users. For example, Google has a record of every search conducted across various devices and every app used on an Android phone, along with interactions made, and location and movement of the phone. If the Chrome browser is used then it knows a user's web history; Gmail, all the emails sent; Calendar, the diary of events; Google Drive, all the files stored; Play, what was bought; YouTube, what was watched; Blogger, what was blogged; and so on. Similarly, Facebook captures a date and time stamped record of every interaction with the platform, including location and device used. It logs every post made, along with edits, comments, likes, and shares; every message made; every photo and video posted; every friend and interaction; movement across websites that have a Facebook 'like' button; information about other apps on the device and any apps connected to a Facebook account. Between them, an enormous amount of detailed personal data is being captured daily for billions of people. By 2017, the Chrome browser has been downloaded over 5 billion times.[21] There were 1.5 billion Gmail accounts[22] and 1.5 billion WhatsApp users in 2018.[23] In 2019, there were 2.5 billion active Android-powered smartphones in use,[24] 2 billion monthly users of YouTube,[25] 1 billion Instagram users[26] and 2.45 billion active Facebook users.[27] That's billions and billions of data points and associated metadata being generated daily.

Much of my life used to be private. My activities, behaviours, thoughts, interests, communications, consumption, work were known to myself, some family and friends, and in a limited way by others such as my bank, employer and government department. What mass datafication means is the creation of rich and detailed data shadows. Consider location and movement. Generally, my position and travel were only known to those I was with or who'd been told of my plans. Now, they are streamed continually via my smartphone, tracked via the GPS sensor and connections to cell masts and Wi-Fi points. My network provider is updated on my location every couple of minutes, as

are any apps that have location tracking enabled. My movement is also tracked via the onboard GPS of my car, the vehicle's automatic passage through tolls facilitated by a transponder, and the scanning of its number plate by ANPR cameras. Passage on public transport is captured by my use of a travel card. Movement on foot can be captured via CCTV, which is increasingly using facial recognition, and by using a MAC-address sensor to track my phone.[28] Many buildings use smart card tracking to monitor and control movement through their spaces. Smart cards are also used to access and pay for public transport.

These location and movement data are not in the public domain, visible to others, as with much social media content. However, they can circulate within data markets. The same is true for other personal data traces. For more than a century, there has been a suite of companies that specialize in gathering together data sets to extract insight and create data products and services. In the digital era, data brokers have grown enormously in number and scale forming a multibillion-dollar industry. As well as large consolidators that purchase, collate and link together data from many different sources, there are more niche brokers specializing in particular data forms, services and markets, such as consumer behaviour, search and background checks, micro-targeted advertising, and credit worthiness. In 2012, Acxiom was reputed to have constructed a databank on 500 million active consumers worldwide, with on average c.1,500 data points per person.[29] By 2018, this had grown to 2.2 billion consumers.[30] By meshing together offline, online and mobile data they claimed to be able to provide a '360-degree view' on consumers with respect to their demographics, financial and purchase history and patterns, political views and values, health, education, address history, criminal and civil case records and so on.

Brokers add and extract value from the data through linking, sorting, matching, analyzing and creating profiles. In other words, they practise dataveillance, monitoring people through their data traces.[31] In turn, they lease/sell the derived data and their data services to a range of clients including public sector bodies, banks and financial services, insurance companies, media conglomerates, retail chains, healthcare providers, telecommunications industries and others. These services are designed to help clients assess,

socially sort,[32] predictively model and profile, micro-target, nudge, and deliver personalized treatment and recommendations for their customers and citizens in various ways. The aim is to minimize risky transactions by identifying which people to marginalize and exclude (for example denying credit and loans, blocking tenancy, not shortlisting for job hires) and to build customer loyalty by identifying who to value and reward (for example through special offers and personalized services). Such decisions might be based on making inferences about how likely it is a person will make their payments or how reliable a worker they might be, or their projected lifetime value if they remain a loyal custom or employee, or how likely they are to move their custom or labour. In this sense, datafication and dataveillance are key processes underpinning the rise of what has been termed 'surveillance capitalism,'[33] in which extracting value from data is a key driver of profit.

For government departments and public sector bodies datafication and dataveillance have shifted the nature of their work and the mode of governance they enact. Since the birth of the digital age in the 1950s, governments have used computers and databases to manage populations and deliver services. Cybernetic thinking in the 1960s underpinned initial attempts to plan infrastructure and services using computer models. In the 1980s and 1990s, personal computers started to become commonplace in central and local government, used for administration and operational delivery, and technologies designed to regulate behaviour, such as traffic management systems, were deployed. The rise of the internet in the 1990 and 2000s, led to large investments in e-government (the delivery of services and interfacing with the public via digital channels) and e-governance (managing citizen activity using digital tools). Our data were captured by governments intent on effectively and efficiently regulating, policing and planning for society, making sure those that deserve and qualify receive payments and services, and countering fraud. The use of digital technologies to mediate governance has deepened in the 2010s with the rise of big data systems. Data brokers provide some of these data and data services. Indeed, a key driver of government becoming data-driven has been a move to make government act more like

a business, and to work more closely with businesses through outsourcing and public-private partnerships. In many cases, this has included privatizing services, and in doing so privatizing publicly owned data, which can then be monetized in various ways (which also exempt them from being released as open data).

Whether the data was intended for the purpose, or whether we like it or not, our data are corralled and combined, value extracted, and used by businesses and government to make all kinds of decisions about our lives. In this way, data are not just left in our wake, but also precede us by influencing how we are treated in subsequent transactions. One of the prices of enjoying the benefits of living digital lives – increased choice, better products and services, improved productivity, efficiency, safety, security, interaction, convenience and so on – is mass datafication and dataveillance. Presently, almost every activity seems to produce data, even many that seem to still be pretty dumb (as opposed to smart). For example, turning on a water tap appears to be an analogue activity, but in the background the utility is digitally networked and the infrastructure and its users monitored and controlled by sensors, transponders, actuators and meters that feedback digital data concerning consumption, pressure, water quality and supply levels to centralized control rooms and databases. When I was growing up we lived small data lives, leaving minor data traces across the public sector and business. Now, we lead lives saturated and thoroughly shaped by data. And we are still coming to terms with this transition and its implications. Living data lives is a work in progress.