**Research Area:**

Big data is a digital phenomenon that enables information to be collected, stored as data, shared, analysed and acted upon at an ever-increasing speed and scale. Combined with the expanse of internet-enabled sensors both in our environments and worn on our person, human activity is being increasingly transformed into data and metadata, which ultimately serves a variety of purposes for many different actors. This fundamentally transforms our understanding of privacy, and muddies the distinction between the public and private spheres. With Big Data comes *dataveillance,* the continuous monitoring of people via (meta)data tracking. Information gathered about us via our online interactions and the Internet of Things is being collected by organisations often without our knowledge or consent, and is used and shared in undesirable and unexpected ways. Dataveillance often draws an apathetic response from surveilled subjects as there is a general sense of powerlessness to enact change. The rewards of submitting to one’s data being collected and exploited are often significant, such as being able to access social media platforms.

The ubiquity of dataveillance has made it virtually impossible to understand the knock-on effects of our actions or inactions. The average person does not have sufficient time and energy to devote to understanding the complexities of how, when, why and by whom data is collected and used. It is often in the interest of data brokers for this to remain opaque, and transparency usually only arises after legislative intervention, such as the EU’s General Data Protection Regulation (GDPR). However, even with such interventions the processes of dataveillance remain shrouded in mystery for most citizens, and there are few informative narratives to aid understanding.

**The goal:**

The goal is to develop social awareness and understanding of dataveillance processes and effects.

I will do this by:

1. writing an accessible narrative that explains what, when, how and where my own data is collected, stored and used and by what actors.
2. producing an accompanying visualisation of the above processes.

**The problem:**

In order to tell such a story, I will need to investigate and be able to understand my own data output. As far as I am aware, there is no singular software solution able to map an individual’s total data output.

**Decomposition:**

**Part 1: Understanding my data output**

1. Identify points of data collection I interact with on a regular basis, for example:

* Google services (search, maps, voice assistant)
* Computer operating system (Windows)
* Social media (Instagram, Whatsapp, Kik, Tumblr)
* Email (outlook, gmail, yahoo mail)
* Phone apps (Spotify, The Guardian, banking app, period tracker, Shazam, ABC listen)
* Phone text messaging/voice call
* Internet of Things (Garmin wearable fitness tracker, google home speaker)
* Opal card

1. Investigate relevant literature alongside data collection/retention policies provided by above services to identify what, when, how and where data is collected, stored and used and by what actors.
2. Record information as spreadsheet.

**Part 2: Visualising my data output**

1. Investigate data visualisation software and find software able to visually represent inputs
2. Being able to express ‘soft’ or descriptive values such as ‘how’, ‘who’ and ‘why’ as data

**Part 3:**

Use Jupyter notebooks

**Pattern Recognition:**

Common data formats, visualizations

**Algorithm design:**