Professor Derek McAuley talks about the [Databox project,](#) the conclusion of a number of years' work in Horizon looking at how third parties may be allowed to process our data in a privacy-preserving way.

What are the issues around companies using the data they collect about us?

Living in a digital world, we all leave digital footprints – data about who we are, where we've been and what we've done. Our phones know where we've been and who we've been talking to. Our heating controllers know when we are at home and how warm or cold we like it to be.

This gives the organisations we interact with a large amount of information about us. We may be comfortable with a company knowing some things about us – such as how many cups of tea we drink in a day. However we should not be comfortable handing over information such as our second-by-second energy profiles from which others could determine when we are at home and when we are not.

Often we end up sharing a huge amount of our personal data with a company, when what they actually want, and what is of value to them, is only a very small statistic derived from that data. Current architectures promote this sharing of the raw data with the service provider so, in addition to relevant data being handed over, other information is also passed on. This is where privacy concerns arise and people have become distrustful.

[What is Databox,](#) and how does it overcome these problems?

Databox is an open-source personal networked device that, rather than having to share personal data with a company, allows the company to run an application on our data on our own computer. By providing this different implementation and a measure of auditability, we are aiming to recover user trust and, in addition, unleash some of the creativity around the use of personal data.

With Databox you never have to give your data away. It stays in your Databox safe and secure, allowing control over exactly what data is shared and with whom. Databox enables you to use your data, and allows other people to use your data, without ever giving it away.

How will users benefit from Databox?

For companies, having to comply with data protection regulation in addition to other sectorial business legislation, Databox technology provides a valuable service in minimising risk associated with accumulating personal data outside of what they actually require. Many companies do not want personal data for other purposes - their business model does not depend on selling the data to somebody else. They just want to provide their customers with a tailored service, to meet their needs and retain their good reputation.

From a societal point of view, companies may be able to offer insight encapsulated in software that enables them to run and process certain data that would be difficult for them to gain access to otherwise, whether due to regulation or issues of privacy. This could prove to be advantageous to the individual - for instance obtaining personalised recommendations provided in complete privacy.

Understanding how data can be exploited is important to both companies and the general public. It is very difficult to detect what an application is doing when connected to the mobile phone network, and so it could have one behaviour on the Wifi where it's monitored and could change its behaviour when it goes somewhere else, with the risk of data leakage. Databox monitors all activity - any information that gets shipped out of the Databox is logged. Part of our role will be to produce simple visualisations to help users understand this.
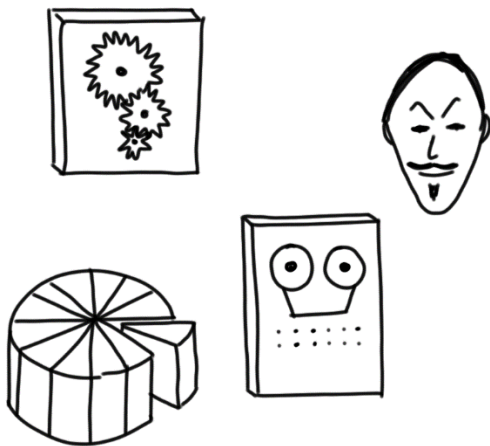
At what stage is the Databox project now?

We are looking at many of the so-called big data problems, where today people try to gather as much information together in one place so they can run their analytics on it. Instead we need to understand the classes of problem we can solve by leaving the data where it is and distributing the analysis over the wide area network - how to get bits of the analytics running in each home so we process the data there and only share the derived information once it's been compressed down to a small number of useful statistics. This is taking a very different approach to the current model that most people have about big data.

Databox has launched the open source software to start building a community who are interested in helping us build the platform - a generic infrastructure similar to the "operating system" of Windows, an environment in which anyone can write applications to run on Databox. However, in the same way that when you buy a new device you expect it to come with certain built-in applications, for instance email and text messaging, Databox will come with standard setup applications, with others available to download, and possibly buy, through the familiar "app store" model.

http://www.databoxproject.uk/