Where does your data go? Mapping the data flow of Nest

mastersofmedia.hum.uva.nl/blog/2017/10/25/where-does-your-data-go-mapping-the-data-flow-of-nest/

October 25, 2017

Introduction

With this research project, we set out to investigate the volume of data created and shared from smart home devices.

Intrigued by the claim that our homes are getting smarter and more connected, we aimed to find out where the data and what type of data of such connected devices end up. We chose to use Nest, a leading company in the smart homes sector, as a means to investigate the data flows of connected devices. Google bought Nest Labs for $3.2bn in 2014. It is likely that Google's motivations for the purchase lay not with the business of home automation as much as the data that these smart home devices, entangled with the growing Internet of Things, have the potential to collect. IBM's CEO Ginni Rometty has labelled big data as "the new oil" (Deutscher), and Google is a platform that knows this like no other.

In October 2011, Nest Labs introduced their smart, a self-learning thermostat connected to the Internet that improves climate control of homes and businesses to save energy. The user has to normalize the thermostat in order to provide his personal reference data set. Nest studies the timetable of this user and learns, for instance, his preferred house temperature. Using integral sensors and mobile geotracking the Nest moves into energy saving mode when it comprehends that the user is not home. Nest also produces the Nest Protect smoke and carbon monoxide detector, the Nest Cam with night vision, two-way talk, sound and motion alerts, as well as optional Nest Aware cloud services. The Nest App brings together all devices, potentially allowing access to the Thermostat, Cam, Protect and other devices in a user's pocket.

On the Nest website we found 116 other third-party devices that work with Nest (Works With Nest). They range from baby cams to smart fridges and light bulbs. These devices track numerous data points from users, including device usage metrics, IP addresses, contact and payments details, and more. This study aims to map the possible data flows between these devices, and to show which companies and entities can potentially make use of the user information from Nest activities. We also propose a creating a public website to highlight this information flow to Nest users.

It is interesting to examine how much data is gathered by all these devices and what companies have access to it. We will look into the specific data that the Nest collects, the details of which may not be clear to consumers that extensive personal information is assembled and potentially shared with third parties. We focused on smart home technology. This technology consists of "applications like security systems and remote monitoring that adapt to a user's presence and habits" (Zeng 2).

Relevance: The internet of things

The Internet of Things, defined as a global Internet-based technical architecture is the concept of connecting any device to the Internet to provide smarter insights on their usage. This can include a large number of appliances, for example, fridges, coffee machines, lamps and thermostats (Morgan). The main focus of the Internet of Things is that everything that can connect, will be connected. To illustrate, the number of devices connected to the Internet by 2021 is estimated to be around 46 billion (Juniper Research). Alongside the increase in devices comes an increase in the volume of data being collected by such technology.

The connection to the World Wide Web allows one to live life just a bit easier, but these devices, often full of sensors and cameras, raise privacy related questions and critique. For example, what happens with the data that is inevitably gathered and with whom is it shared (McDonald)? A debate about the security of these devices is also prevalent (Weber 24). However, this paper will only focus on the debate about sharing of information. We are moving towards a world in which more and more devices and products are connected to the internet, making them remotely accessible from other devices. It is unclear however who receives all the data that is gathered by these devices, and what is done with this data. That leaves us with one important question: who is actually the true owner of that information?

Methodology: Data Gathering

In setting out to analyse the data flows associated with Nest technologies, we identified two relevant strands to our research. The first was the data and information that Nest devices tracked themselves from the user. For this, we were interested in the information captured by the devices themselves.

Secondly, it was necessary to gather information about all the third party integrations and devices that could potentially be connected to Nest devices. To do this, we consulted the Works With Nest website, which provides product descriptions of all 116 devices that can be connected to Nest products (Works With Nest). Using a website scraper, we compiled a spreadsheet of all the names and descriptions of these devices, which we manually tidied and added additional context to, such as detailing the Nest products they related to. As one of the questions we were hoping to answer in this research looked to connect the links in corporate ownership between the manufacturers of the devices sharing and making use of the data, we also conducted research into the parent owners of each device. Once all this information was displayed in the spreadsheet, we were able to proceed to visualising the data flows.

1. Data collected by Nest devices

First, we looked at the information and data utilised by Nest. To find the complete list, we consulted the legal terms and conditions of Nest technologies, and itemised each data point listed on the site (Nest). We made note of every piece of personal information and data that users consented to submitting once they purchased a Nest device and created their online profile and compiled a document with all the information we could gather on these data points. Some of the data relevant to all devices, such as the IP address and mobile location data of

users, comprises of ‘indexical data’, which potentially allows for identification. According to Rob Kitchin, indexical data is important because it “enable(s) large amounts of non-indexical data to be bound together and tracked through shared identifiers, and enable discrimination, combination, disaggregation and re-aggregation, searching and other forms of processing and analysis” (10). Flagging this in the context of security vulnerabilities is important.

Other information was specific to the device, such as smoke and carbon monoxide levels (Nest Protect), and access to video content (Nest Cam). In total, these were the number of data points collected by each Nest device.

Device	Thermostat	Cam	Protect
No. of data points captured	24	25	26

Next, we created graphics illustrating all the information tracked and stored by individual devices, taking the information from obscured legalese on the terms and conditions section of the Nest website and placing it alongside the devices themselves.

A full list of individualised data points per device can be seen here. From here, we proceeded to analyse the data connections between the Nest devices and third party integrations.

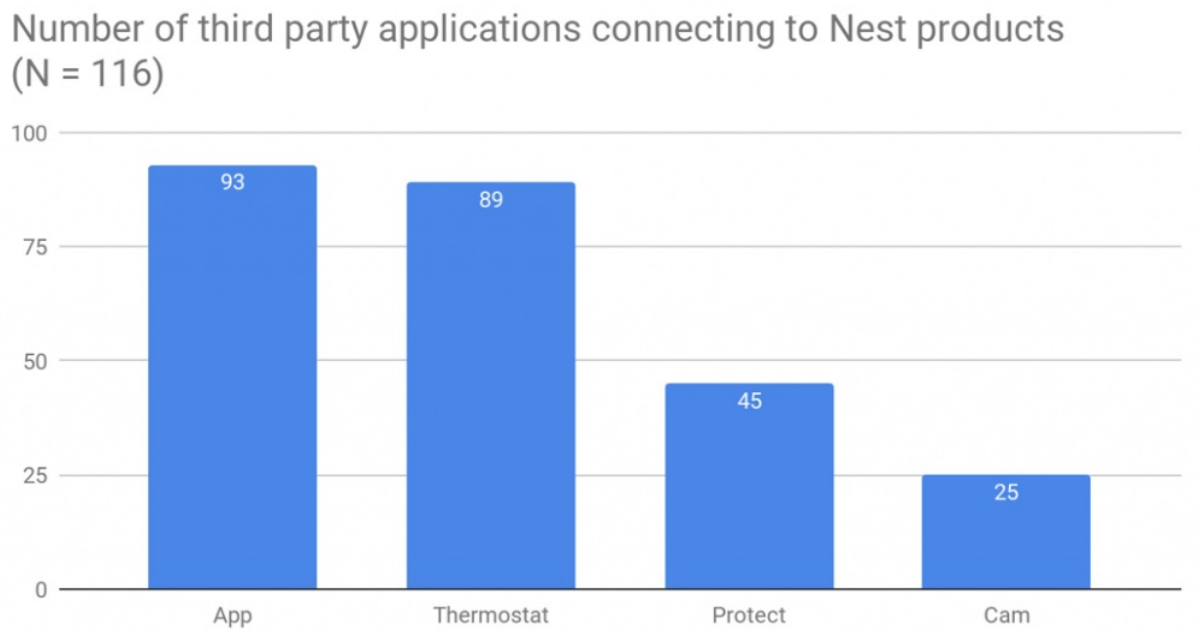
2. Sharing of data with third party integrations and devices

This section will mainly deal with the different visualisations and what they could bring to light. For this, the visualisation is was inspired by Dodge and Kitchin’s visualisation in abstract space, to show how different applications connect to Nest products (42).

A step-by-step guide of the data gathering and visualisation can be found over here. As previously discussed, third parties are able to integrate their products with the Nest using its API. Users give their consent to Nest to share their data to the third party applications they use by accepting the terms and conditions. Our use of mapping techniques was informed by Dodge and Kitchin's view that mapping can be used to "exploit the mind's ability to more readily see complex relationships in images, providing a clear understanding of a phenomenon, reducing search time, and revealing relationships that may otherwise not have been noticed" (2).

To provide a greater insight on the possible data flows that come with integrating third party applications into Nest products, we set out to map the hypothetical flow of data between devices. An important sidenote to this, is that it is not possible to map out the actual data flows, since we are limited to the descriptions of possible data flows.

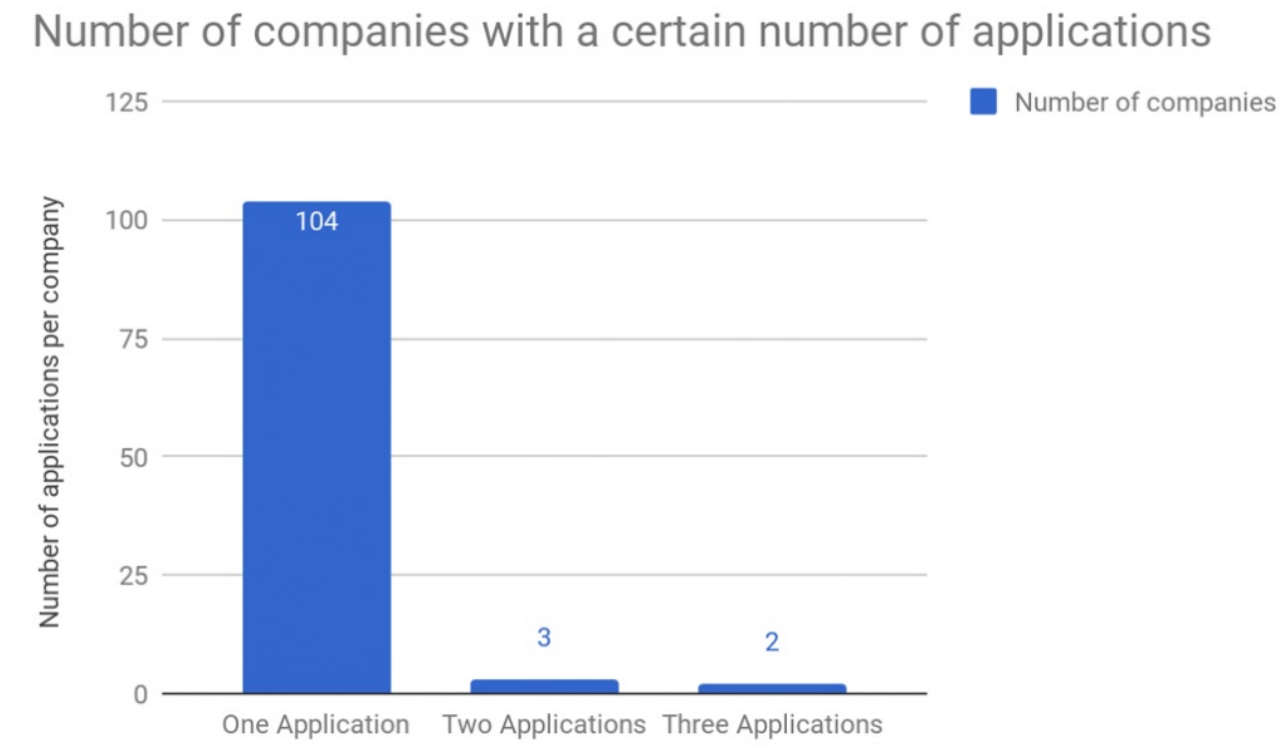
Firstly, an overview was created of the number of different third party applications that connect to each of the Nest applications (Thermostat, App, Cam and Protect). This is because each different Nest product collects different data, even though a lot of it is similar. Nest provides info on what applications use which products and one application can be used with multiple devices. To visualize this, we had to manually go through each webpage of third party applications and note down what each application used. The dataset can be found over here. This resulted in the following graph.



As shown in the graph the Nest app and Thermostat are easily the most connected devices. This can be explained by the fact that the Thermostat and the App were the first Nest products to come out.

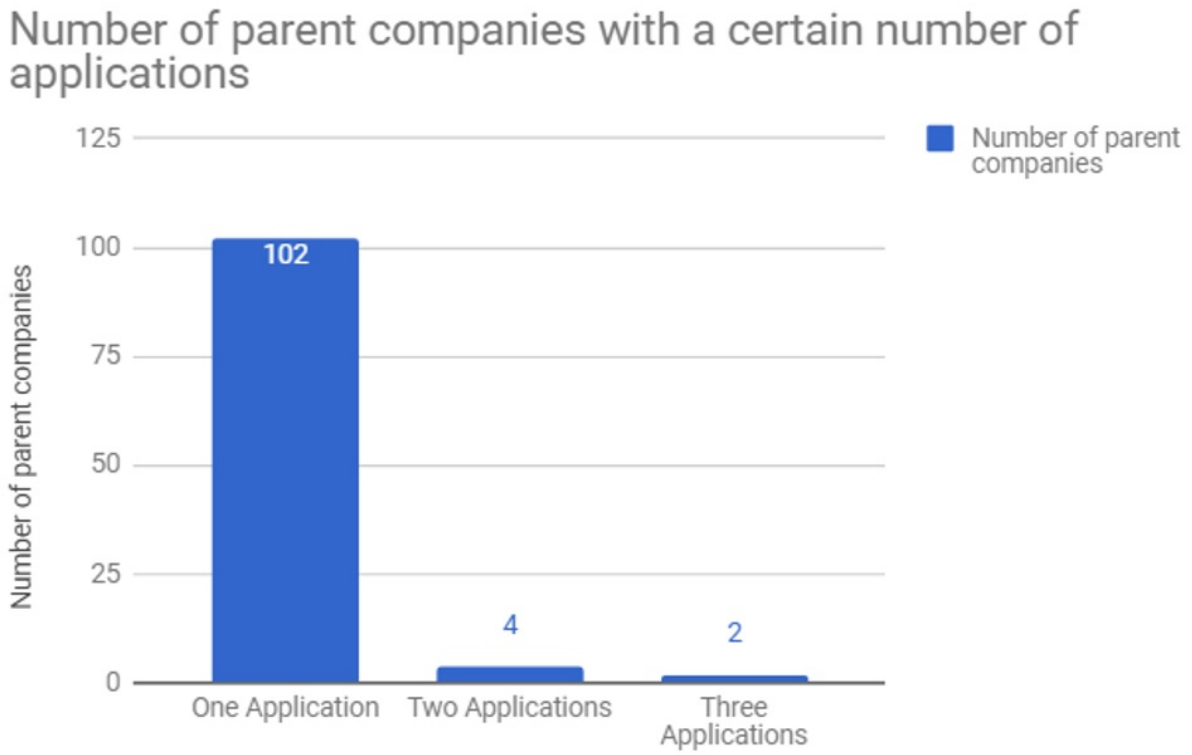
Even though this provides an overview of the number of third party applications connecting to Nest products, this does not necessarily provide an answer to the question of whether these third party applications are owned by a number of large companies, or by a large number of independent companies.

By creating an overview of each different company that owns one of these applications, we were able to assess if some companies have a large number of applications. This provided us with the following graph.



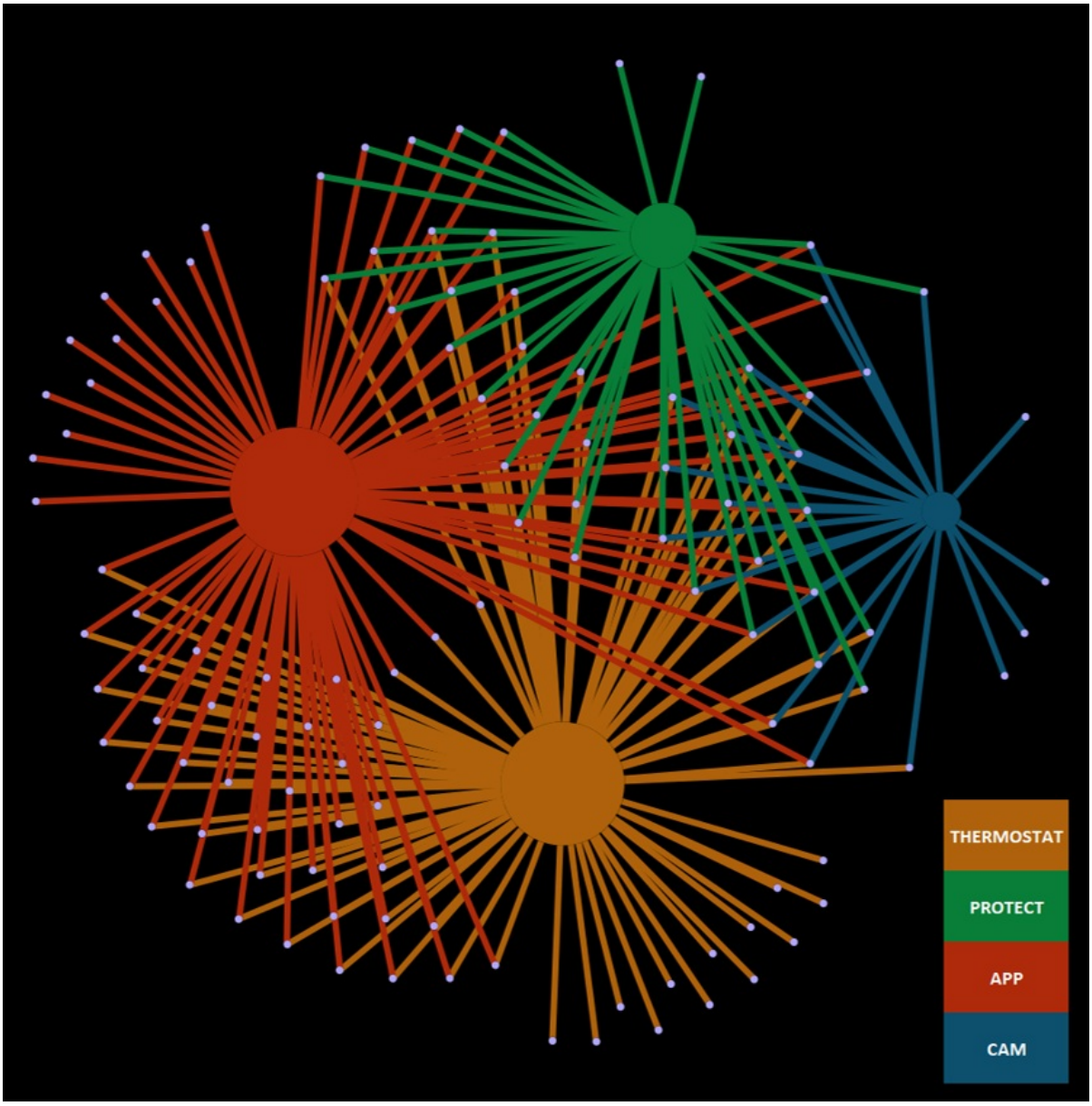
As shown in the graph, only a few of the companies have more than one application, however, no company has more than 3 (Deviate, 2; MaaDoTaa, 2; Roomie Remote, 2; Google, 3; IFTTT, 3). This shows that so far there are no clusters of companies dominating the manufacture of integrations connected to Nest.

However, a large number of companies are owned by parent companies – for example, Nest and Google are owned by Alphabet. Consequently, a number of the companies in the previous graph could be owned by the same company, to assess whether this is true, we created another graph, where each company is replaced by its parent company. This led to the graph below, which interestingly only added one extra company (Whirlpool) that has 2 or more applications, the other companies have either changed ownership.

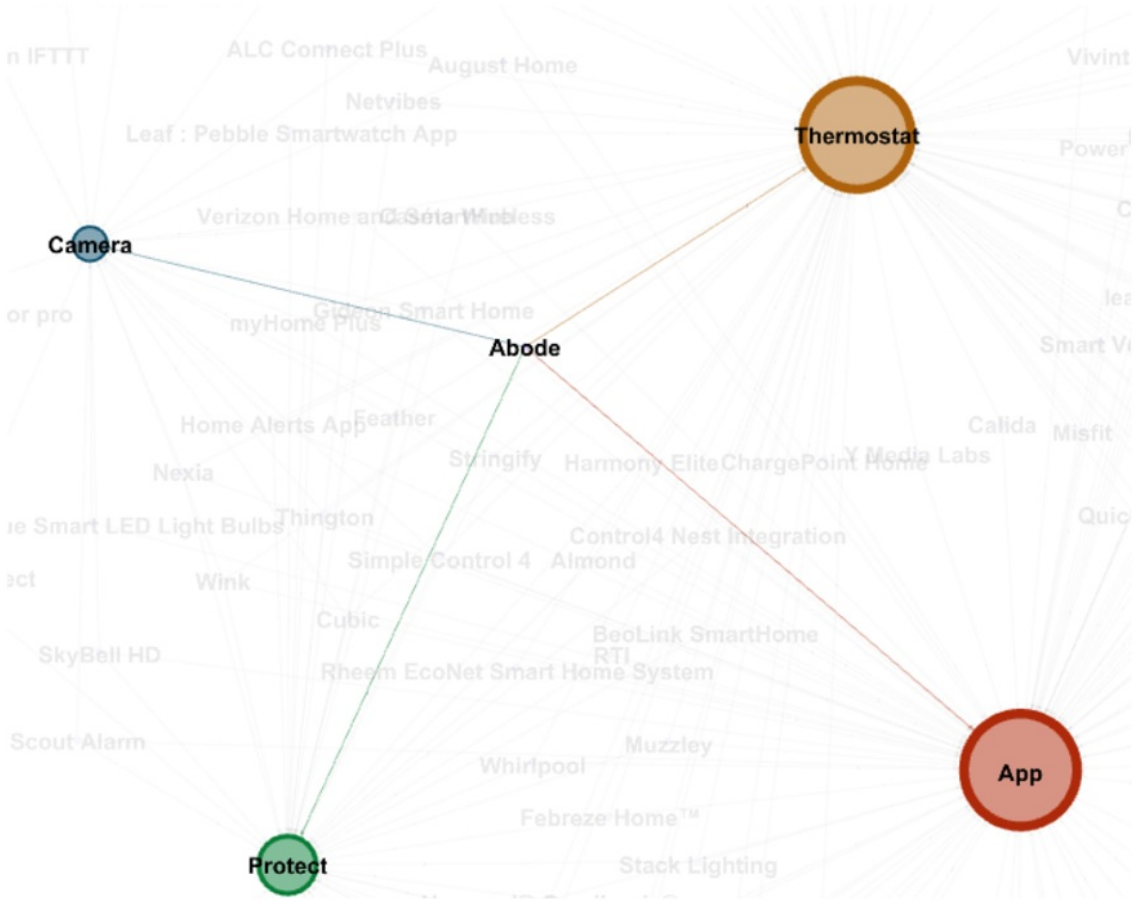


This is interesting because it was possible that there would be a small number of large companies that would have the biggest share in the number of applications. This visualisation however, shows that this is not the case.

Lastly, a visualisation could also be made of the different connection each different application makes with the different Nest products. For example, an application that links with the Cam and the Thermostat, will be clustered towards those two, while being further from the App and Protect. Furthermore, each outgoing edge will be coloured as it's target. This shows the different connections the third party applications make and also shows what data could be shared between devices.



This could be used to further provide insight to consumers to where their data could possibly go. An idea for this would be to provide an app, or a website, where people can fill in the applications they use. After this, they could see to which Nest product the application is connected, after which it is possible to show which data could possibly be shared. An example is shown below.



This way, consumers could be informed about the amount of data they share, consciously or unconsciously, and that could prove to be useful possibly.

Utilising new media: Illustrating data flows for Nest consumers

Armed with this information about the huge volume of potential data collection and sharing via Nest devices, we would propose creating a simple website which Nest customers could use to review exactly what information and data they are potentially making available to Nest and their partners, even at aggregate level. Using menus, the user would be able to select the Nest device(s) that they use in their home, and add on any integrations that they also use. This would create a visualisation using the information we have gathered to show all data points that could be stored and shared by their single device. The goal is to improve visibility around the issue of data sharing from smart home devices, and provide some insight into the nature of data capitalism as practiced by technology companies such as Nest. With the smart homes industry blossoming, this is an area with potentially huge big data privacy concerns. As Nissenbaum and Baracos have argued, informed consent by way of tick box and mass anonymity alone are “ineffective against the novel threats to privacy posed by big data” (32). Arming the consumer to make better decisions about the use of their data is one way of ensuring greater transparency around the issue.

Conclusions

In concluding, we have gained an insight in the type of data that one of the leading companies

in the growing smart home space is collecting and sharing with partners. While their collection of the data is perfectly legitimate, much of the information is largely 'buried' in the legal terms and conditions on their site. By placing this information in clearer terms, we are able to demonstrate the data collection capabilities of each of the Nest devices. We have also demonstrated the potential overall reach of third party sharing of data with other companies, illustrating the connection between the devices and their sharing of data.

Further questions arise as a result of this research. The most relevant relate to the implications that such big data harvesting has on users, and the potential uses of aggregate levels of such information. With all the benefits of connected devices also come risks concerning security and privacy violations with the vulnerability of hacking being one of the most urgent.

Bibliography

Barocas, Solon and Nissenbaum, Helen. 'Big data's end run around procedural privacy protections'. *Communications of the ACM* 57.11 (November 2014): 31-33.

Deutscher, M. "IMB's CEO Says Big Data is like Oil, Enterprises Need Help Extracting the Value", *Silicon Angle*, 11 Mar.2013. <https://siliconangle.com/blog/2013/03/11/ibms-ceo-says-big-data-is-like-oil-enterprises-need-help-extracting-the-value/>. Accessed 25 Oct. 2017

Dodge, Martin and Kitchin, Rob. *Atlas of Cyberspace.* London: Pearson Education Ltd, 2001.

Gallagher, S. "The future is the Internet of Things – deal with it." *The Times*, 29 Oct. 2015. <https://arstechnica.com/features/2015/10/the-future-is-the-internet-of-things-deal-with-it/> Accessed on 25 Oct. 2017.

Juniper Research. *'Internet of Things' Connected Devices to Triple by 2021.* <https://www.juniperresearch.com/press/press-releases/%E2%80%98internet-of-things%E2%80%99-connected-devices-to-triple-b> Accessed 2 Oct. 2017.

Kitchin, Rob. *The Data Revolution. Big Data, Open Data, Data Infrastructures and Their Consequences.* Los Angeles: Sage Publications, 2014.

McDonald, Aleecia. "Laws Can Ensure Privacy in the Internet of Things". *The New York Times.* Ed. Dean Baquet. 2013. 2 October 2017. <https://www.nytimes.com/roomfordebate/2013/09/08/privacy-and-the-internet-of-things/laws-can-ensure-privacy-in-the-internet-of-things>

Privacy Statement for Nest Products and Services. 2017. Nest. 23 October 2017. <https://nest.com/legal/privacy-statement-for-nest-products-and-services/>

Weber, Rolf H., and Romana Weber. *Internet of Things*. Vol. 12, Springer, 2010. *Google Scholar*, <http://link.springer.com/content/pdf/10.1007/978-3-642-11710-7.pdf>

Works With Nest. 2017. Nest. 22 October 2017. <https://workswith.nest.com/>

Zeng E., Mare S., Roesner F. “End user security & Privacy concerns with Smart Homes” *Symposium on Usable Privacy and Security (SOUPS)*. 2017.

Thirteenth Symposium on Usable Privacy and Security ({SOUPS} 2017), July 12–14, 2017 • Santa Clara, CA, USA https://www.usenix.org/system/files/conference/soups2017/soups2017-zeng.pdf. Accessed on 25 October 2017.