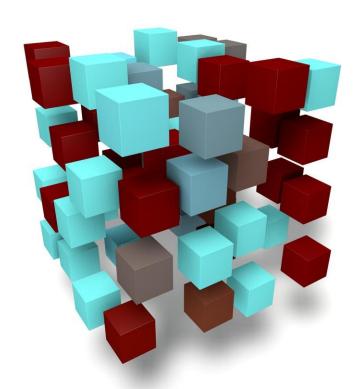


Market engagement in the Databox project



Summary Report

Results of the stakeholder consultation

PREPARED ON BEHALF OF HORIZON DIGITAL ECONOMY INSTITUTE
October 2017



Report commissioned by:

Horizon Digital Research Institute

http://www.horizon.ac.uk/

Report authors:

Mattia Fosci, Rob Johnson www.research-consulting.com

Contact:

rob.johnson@research-consulting.com

Report dated: October 2017

© The University of Nottingham



Executive Summary

This document looks at market engagement with the Databox concept, as it emerges from a limited number of conversations with project stakeholders, business consultants and industry experts. It seeks to explore its value proposition for individuals, 'data-rich' companies and 'data-poor' companies.

Hypothesis	Stakeholder perspective
For individuals, the main value is the ability to control and potentially monetise third-party access to their personal data.	Interviewees are sceptical about market demand for more control over personal data. Concerns were raised over a general lack of awareness of the problem, and the difficulty in deciding what data should be shared. The idea of creating a data market whereby consumers sell personal information to companies is deemed interesting in theory. However, doubts were raised over the economic viability of the concept, as the data market is currently very competitive.
For data-poor companies, the main value is in the additional insights that could be gained from the Databox	Our interviewees agreed that data-poor companies might in principle be interested in paying to access the Databox insights. Concerns centre on the accuracy of such insights, as well as their cost. Moreover, the Databox is only likely provide valuable information if it achieved a high level of market penetration.
For data-rich companies, the main value is in reducing the legal, economic and reputational risks connected to new data protection regulations.	Data-rich companies are indeed worried about the impact of the GDPR and other data regulations. However, it is not clear to what extent the Databox could reduce their exposure while providing market insights comparable to, or in excess of, what they currently have.

Recommendations to advance the development of the product focused on three areas:

- 1. clarifying and simplifying the user interface and user experience for individuals;
- 2. clarifying the technical aspects for companies to **generate 'trust' in the data**;
- 3. reviewing the **business model** by looking at production costs, revenue streams and revenue forecasts.

A trial with consumers could provide more information into each of the above areas, and either validate the current vision for the product or suggest **different applications**. Examples of alternative applications for the Databox suggested in our consultation include:

• as a polling machine to get market research data from a small sample of consumers;



- as embedded code that surfaces data privacy for all applications and standardises consent; or
- as a secure data access point that manages access to personal data stored in the cloud.

These examples show that the concept of the Databox is adaptable, and the final shape of the Databox can be best determined by **direct engagement with consumers, companies and regulators**.



Contents

Execu	itive Summary	3
Conte	ents	5
1.	Introduction	6
1.1	Scope of work	6
1.2	Methodology	6
1.3	Limitations	6
2.	Hypotheses	8
2.1	What are the customers' profiles?	8
2.2	What is the value created by Databox for customers?	9
3.	Insights from industry experts	10
3.1	Who would be the customer?	10
3.2	The value proposition for individuals	11
3.3	The value proposition for data-poor companies	12
3.4	The value proposition for data-rich companies	13
4.	Recommendations	14
4.1	Focus on the user experience (for individuals)	14
4.2	Clarify technical features (for companies)	15
4.3	Review the business model (for investors)	16
5.	Alternative models and possible ways forward	17
5.1	Alternative models	17
5.2	Next steps	17
5.3	Conclusions	18
5	5.3.1 Value proposition for companies	18
5	5.3.2 Value proposition for individuals	19



1. Introduction

1.1 Scope of work

Research Consulting was asked by the Horizon Digital Research Institute at the University of Nottingham to investigate the potential for market engagement with the Databox, an innovative technology developed by Horizon researchers that provides localised storage of, processing and access control to personal information. Our work focused on gathering evidence to help steer the Databox towards product/market fit:

- who might use the Databox (users) and who might pay for it (customers);
- what the value proposition is for both users and customers;
- what practical applications and markets have the greatest potential; and
- what the potential business models might be.

1.2 Methodology

We gathered qualitative data through desk-based research and interviews with industry experts from four companies:

- An SME operating in the telecommunications, security and Internet of Things (IoT) markets;
- A global cloud platform and data analytics company;
- A multinational telecommunications company;
- A global consumer goods company.

Moreover, we used *Value Proposition Design* (see Figure 1 on following page) as the underlying methodology to analyse the value of the Databox for customers. This methodology encompasses two questions:

- (1) what problems the technology is solving (Customer Profile canvas), and
- (2) whether customers are aware of these problems and willing to address them, which in turn determines the perceived value of the solution (Value Proposition canvas).

The remainder of this document summarises the key reactions, concerns and suggestions voiced by industry experts and suggests options for moving the technology forward. Anonymised quotes from our interviewees are included throughout the document to illustrate the views and opinions expressed.

1.3 Limitations

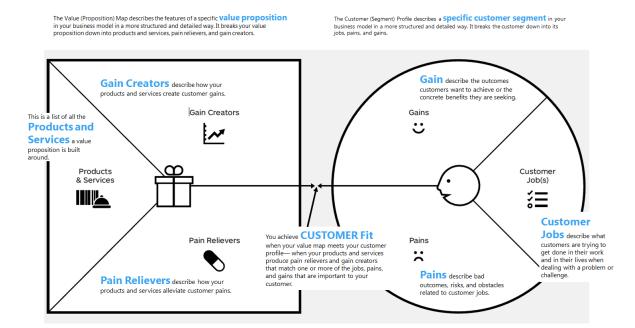
This study is based on the views of a small sample of four industry experts working in the areas of Internet of Things software and hardware, data, information technology and communications. Experts shared their personal views, which do not represent the position of the organisation they work for. As such, their views cannot be considered representative of the whole ecosystem of potential business partners or customers. Instead, they should be considered preliminary insights into the market challenges and opportunities facing the Databox.



Considerations over the business case for the Databox were also complemented by the views of Dr Nick Gostick, project associate. Dr Gostick has extensive experience creating, supporting and managing technology businesses at the Universities of Sheffield, Warwick and Birmingham. His insights were therefore used to inform the conversations we had with industry experts.

Finally, the study is also based on Research Consulting's understanding of the Databox, as it results from conversations with the lead researchers – Professor Derek McAuley, Dr Richard Mortier and Dr Hamed Haddadi. Although our team's limited technical knowledge did not affect our study on market engagement with the Databox, it has limited our ability to answer some of the technical questions asked by the industry experts. Some of these questions are therefore reflected in this document in the form of recommended clarifications for the research team.

Fig. 1 – The Customer Profile and Value Proposition canvas (Alexandre Osterwalder et al.)^a



^a Osterrwalder A., Pigneur Y., Bernarda G., Smith A., Value Proposition Design: How to create products and services customers want, Wiley, 2014



2. Hypotheses

2.1 What are the customers' profiles?

Using the conceptual framework of Value Proposition Design, this section presents a set of hypotheses concerning the problems solved by the Databox, the profile of individuals and organisations facing those problems, and the perceived value of the Databox as a solution to these problems. These hypotheses are based on the initial conversations we had with Databox researchers, as well as discussions with our project associate and industry experts. We identified three types of customers:

- Individuals concerned with protecting their privacy
- 'Data-poor' companies lacking direct access to their customers' data
- 'Data-rich' companies concerned with the impact of the General Data Protection Regulation

Using the customer problems canvas, the tables below describe some of the research hypothesis actions or tasks each customer is undertaking that would be affected by the Databox (called 'jobs'), the obstacles or problems customers face when performing such actions/tasks ('pains') and the positive outcomes or concrete benefits sought by customers ('gains'). The validity of these hypotheses will then be analysed more in detail in section 3.

Table 1 - Profiling individuals (hypotheses)

Jobs	 Access content and services online and offline Complete transactions online and offline Use IoT-ready home appliances
Pains	 Privacy breach Online fraud and identity theft Lack of control over data/sense of being 'spied on'
Gains	 Ability to determine what information is shared (privacy) Ability to get lifestyle insights (e.g. own energy use) Potential to monetise transfer of personal data (e.g. by receiving vouchers or discounts from companies)

Table 2 - Profiling data-poor companies' (hypotheses)

Jobs	Acquire data about their customers to improve marketing, product development and revenue
Pains	 Lack direct access to customer data (e.g. because they sell cheap products in physical shops) Purchase expensive data or market research services that puts them at a competitive disadvantage vis-à-vis data-rich companies Have incomplete or inaccurate picture of their market



Gains	Ability to gain better insights into customers' habits
	 Ability to acquire data about the competition
	 Save money by acquiring data directly from the customer

Table 3 - Profiling data-rich companies (hypotheses)

Jobs	Acquire data about their customers to improve marketing, product development and revenue
Pains	 Legal risks caused by GDPR provisions and interpretation (especially concerning the storage and processing of "sensitive data") Increased costs to improve data security Reputational risks (collection and loss of data)
Gains	 Reducing costs and legal risks while keeping access to data Increased granularity of data acquired from users Retain customer loyalty by engaging directly in data transactions

[&]quot;If web and IoT technologies are to grow and evolve, it is absolutely valuable that people have confidence about the way their data are collected, protected and used."

2.2 What is the value created by Databox for customers?

It is assumed that the Databox would address many of the above problems, and that it would therefore have clear value for customers. The table below summarises the hypothetical value proposition of the Databox.

Table 4 – The Databox' value proposition (summary, for all actors)

Key Databox features	 Local storage of personal data Local processing of personal data Access control to personal data (demonstrates explicit consent)
Pain relievers	 Ability to control personal data Local storage and processing of data reduces legal risk for companies Increase data security (the data is not 'online by default') Reduced security concerns for companies (raw data is not stored in the cloud)
Gain creators	 Ability to monetise personal data (through vouchers, discounts etc) Direct access to high-quality personal data for data-poor companies Ability to combine personal data from multiple sources gives richer picture of customers Retain customer loyalty by engaging directly in data transactions



3. Insights from industry experts

This section summarises the perspective of the industry experts interviewed. We use a colour-coded system to show the level of agreement/disagreement with a particular hypothesis.

- Green = the hypothesis is currently true and verifiable
- Amber = the hypothesis is currently partially true or untrue, but prospects may change in the future
- Red = the hypothesis is untrue and there are no prospects of this changing

3.1 Who would be the customer?

Customer type	Interviewees' position	Agree/ disagree
Individuals	Most interviewees agreed that it is more likely for individuals rather than companies to be paying customers for the services provided by the Databox. Companies might act as Databox partners (e.g. by embedding it in one of their products) if there was customer demand for increased data privacy and security. The consensus was that this is not yet the case, but companies are monitoring market awareness around this issue.	
Data-poor companies	Data-poor companies might be interested to pay for the data or insights provided by the Databox as a way of bridging their competitive gap, but are unlikely to pay for the Databox itself. Moreover, the price paid for such data or insights is likely to be very low unless there is a demonstrable impact on the bottom line (acquisition costs < lifetime value of the customer).	
Data-rich companies	Despite concerns about the GDPR, these companies are unlikely to pay for data that they can currently gather for free. It is however possible that some of these companies could partner with the Databox (e.g. providing the hardware).	

"How much do people care about their personal data and privacy? Some people are taking steps to protect their privacy, but the majority of the population is oblivious"



3.2 The value proposition for individuals

Customer type	Interviewees' position	Agree/ disagree
Consumers want to protect personal information	Whilst interviewees agree that consumers are concerned about the security of their data, it was felt that public perception around data security would not increase by storing data locally. Anecdotal evidence points to the fact that people consider cloud services safer than their home machines, which are vulnerable to being hacked, broken, or stolen.	
Control what information is shared (privacy)	Currently there seems to be willingness to share personal information with companies, although the public's attitude may change in the future. However, difficulties in determining what data should be shared with third parties, and the time-consuming nature of the activity, is likely to limit engagement to the niche of people that are familiar with the subject.	
Get lifestyle insights	The ability to get information about personal/household lifestyle would be valued, especially if combined with recommendations on how to save money, what products to buy and so forth. However, this information would have to be presented simply and conveniently (e.g. on a mobile phone).	
Monetise transfer of personal data	Interviewees thought that providing vouchers or discounts would be the most effective way to incentivise individuals to share their personal data through the Databox. However, they also noted that most people would find it difficult to determine what data should be shared and that incentives would have to be 'significant' for users to engage.	
Would individuals pay for the Databox?	At this stage, it is unlikely that individuals would pay for the Databox, but things may change in the future. It is clear, however, that only a minority of customers would be prepared to pay and only relatively small sums. Noting that 'people do not have want to pay for internet routers', one interviewee suggested that the Databox used a service model (e.g. monthly subscription) rather than a product model (upfront cost).	

"You do not want a service to request you to make decisions every second. Financial services make several queries per second."



3.3 The value proposition for data-poor companies

Customer type	Interviewees' position	Agree/ disagree
Gain better insights into customers' habits	Data-poor companies are businesses that lack effective and scalable mechanisms to collect data from their customers, for instance because they sell low-value products in shops. They are likely to value the data provided by the Databox as a cheaper and more direct way of acquiring higher quality data about their customers. Current approaches with big data are inference-based and usually they are wrong: the value of a databox is that it takes your data and allows you to be more specific about what one user really wants.	
Gain a more complete and accurate picture of their market	Databox could also provide data about a consumer's purchase of competitor products. However, interviewees seemed less sure about its ability to provide market-level insights. In particular, doubts were voiced with regards to the Databox's data processing capabilities and data consistency (see below).	
Save money by acquiring data directly from the customer	The data market is already competitive, so the cost of data acquisition is not perceived as a pressing problem – and Databox is not perceived as a cheaper alternative. Perceptions may change once the implications of the GDPR requirements become clearer, but for now the focus is on data quality rather than cost.	
Would data- poor companies pay for the Databox?	Data-poor companies would not pay for the Databox, and would not be the ideal partners in a joint venture. They would solely be interested in accessing consumer data through the Databox, and could consider paying small sums for this data. Because they hold limited data, their worries about GDPR are less intense compared to other companies.	

[&]quot;The main interest from of a company perspective would be to gain direct access to customer information that is currently held by other companies"



3.4 The value proposition for data-rich companies

Customer type	Interviewees' position	Agree/ disagree
Achieve GDPR compliance	Interviewees agreed that the GDPR is a big source of concern for all companies and that anything that can make it easier for companies to comply with GDPR is valuable. However, it was not clear to what extent the Databox would reduce company risk/costs. Interviewees presumed (perhaps incorrectly) that any information that is remotely queried by companies would go back into the cloud, meaning companies might still be legally responsible for the data they get from the Databox. The Databox could increase transparency, but to achieve that it would have to make available extensive metadata, including details on the approach taken to data-processing.	
Acquire better insights	Interviewees were sceptical about the Databox's ability to provide better insights than is currently the case. On the one hand, they noted its ability to combine multiple data-sources from the same customer. On the other hand, they stressed that data processing would be very resource-intensive in order to answer the many different queries companies make (given the high frequency and different nature of the queries).	
Limit reputational risks	Databox is likely to have the strongest positive impact in terms of reduced reputational risks for companies. Perception-wise, the idea that personal data are stored locally and that companies only have control over anonymised insights, or data which users have explicitly consented to, reduces reputational risks for companies even in the event of data breach. However, this depend on consumers showing increased awareness about this problem.	
Would data-rich companies pay for the Databox?	Databox would be hugely disruptive to the big-data market: having invested heavily in the model they have, it is doubtful that big companies would be willing to transition to a localised data processing model. Should this become the best option due to, e.g. regulatory changes, big data companies might decide to develop their own system – where they retain access to and control over the data.	

"Different providers will want the data processed in different ways. The logic in the Databox is going to be very hard as it is going to be very difficult to process the data in all the different ways customers would want."



4. Recommendations

This section summarises the suggestions and recommendations put forward by industry experts during interviews. These have cropped up organically in the conversation, generally as part of the conversation around the barriers to adoption faced by the Databox.

The suggestions have been explored in some detail through follow-up questions, and then organised based on whether they focused on improving the product for individuals, clarifying the offer for companies, or present a convincing business model for investors.

4.1 Focus on the user experience (for individuals)

Make Databox easy to use	Make the Databox interface easy to use by making it run on smartphones. For instance, Android provides some of the capabilities that tell the users what data they are collecting and using — how can the Databox successfully integrate with the existing user experience of Android users?
Make the data "physical"	If customers have to pay for the Databox (whether using the product or the service model), the Databox should be a physical product – this will help make the question of data privacy 'real' for consumers. This may be a better option for data-conscious consumers. Costs would have to be limited, however.
or embed within existing home products	If customers pay for the Databox as part of an existing service subscription, or do not pay at all, then integrate the Databox in existing home hardware. The concept of the connected home is growing and likely to generate data privacy concerns. Databox could be part of the home hub that controls all appliances. This would make it more commercially viable and accessible by less dataconscious consumers.

"Databox will be a commercial product if the consumer market goes that way - if it happens it will be because users demand it."



4.2 Clarify technical features (for companies)

Address concerns around software security	Can you demonstrate that the data in the Databox is better protected than in the cloud? The process of companies asking for data is run by software, so you can have rogue companies polling for the data: how do we make sure that the process does not get compromised?
Address concerns around physical security	Would a physical box be more secure than the cloud? What happens if the hardware is damaged, broken or gets stolen? How do I recover data or prevent others from accessing it?
Clarify process	How do companies know what is in the Databox without interrogating it? Would they have to poll each Databox individually to have access to the data?
Clarify what data it would provide	What data is held in the box and in what format? Is the data immutable or mutable? Is it supplied with an authentication or audit trail that would allow you to say that it was 'true'?
Build in transparency to generate trust	The Databox would need to communicate not only what data it has, but also how it has obtained the insight, which is important for privacy-supporting audits. It would have to be fully open about all the standards it uses. It is important that the Databox enables companies to check that the processing was 'true'. Companies will need details about the integrity of the data and of the processing in order to trust that Databox is properly running the code they send.
Build identify verification processes	Data quality and credibility is key. Companies insisted that it is especially important for the Databox to 'prove identities' – strong and auditable processes to verify users' identity if and when needed.
Optimise data transfers	Note that broadband networks are generally designed for fast upload and slow download. Multiple remote requests for data (in a hypothetical situation where the Databox has market penetration) are likely to strain current capacity. An assessment of network capacity and efforts to optimise data transfers are needed to ensure fast response.

"Companies will want the Databox to be fully open about all of the standards it uses: they will want to know what information it has but also how it has obtained it. The balancing act is between protecting the IP and trusting the data."



4.3 Review the business model (for investors)

Market penetration	Consider the level of adoption of the technology: how would Databox work when you have limited market penetration? Would companies still need to hold the data for those users that do not have the databox?
Work out production costs	What are the costs associated with long-term data storage? What are the costs to give the Databox sufficient processing power and memory to perform increasingly complex data processing? Concerns are that if the Databox stores encrypted data it will need high processing power to deal with frequent, multiple requests from companies – so the manufacturing cost could go up to hundreds of pounds.
Study your customers	Respondents considered the market immature and stated that customers would not pay to solve a problem they don't yet know they have. Customers can be incentivised by companies to give control over their data, but such incentives would have to be significant and the decision would have to be simple (one click) and infrequent (e.g. once a week).
Companies are difficult partners	Interviewees doubted that companies would pay for the Databox at this stage. It is more likely that companies would be willing to be 'partners' once there is demonstrable demand from their users: under what conditions would companies happily surrender control over their data to the Databox? Would they be happy for other companies to be able to access their own data? Interviewees suggested that companies would not part with their data unless mandated by regulation or incentivised by strong market demand, and that they would not want competitors to access their data (but would be interested to access competitors' data – unsurprisingly!).
Data market	Would companies pay to access individual data, and how much would they pay? The Databox would need to build a micro-payment application (which is not cheap) - and by buying individual data companies are likely to raise their transaction costs, which would in turn raise the price of items. Interviewees suggested that the system would work with high value purchases only, but not with large scale consumer products.
Articulate the technological USP	Some interviewees saw the Databox as a customer interaction solution, rather than a technological solution. It would be beneficial to articulate the technological 'unique selling point' of the Databox in simple terms for non-technical audiences, as well as its benefits compared to alternative approaches such as Blockstack.

"Customers don't even want to pay for broadband routers – a service model would be more effective than a product model"



5. Alternative models and possible ways forward

This section presents some alternative approaches put forward by the interviewees for the commercial development of the Databox. It also explores some possible next steps for the research team.

5.1 Alternative models

Polling machine	The Databox could be used as a polling instrument to get accurate statistical data, cheaply and effectively. It could be distributed to relatively few customers (e.g. a few thousands) who would be incentivised to use it. This is a similar model to audience response systems. This way, Databox would provide an alternative (and GDPR-compliant) market research tool that does not rely on extensive data mining.
Embedded code	Another model is to embed the Databox into data-intensive apps. The Databox software would run data processing operations for each app on the local machine (smartphone) and incorporate familiar interfaces to give companies the permission to access the data. For the user, this would help surface data privacy for all applications and standardise the way they give and remove consent to data sharing across services.
Secure data access point	The Databox could offer a highly reliable and secure access management system for data stored in the cloud. The expert suggested that it is highly unlikely that all personal data can be stored in the Databox (and, incidentally, that the Databox would face storage limitation problems over time). Instead, the data could be stored and processed on cloud servers managed by the data controllers, but the Databox could be the secure proxy that manages access to the data.

5.2 **Next steps**

Protect IP	Interviewees suggested that the Databox team look at whether the technology, or part of it, is patentable. As market conditions do not yet appear to be mature, strong intellectual property protection would place the team in a strong position if and when conditions change. For instance, large data companies may look at acquiring the technology once the implications of the GDPR become clearer and/or consumer attitude changes.
Trial the Databox	The commercial viability of the Databox is based upon three conditions: (1) that the product can provide valuable information to companies, (2) that consumers show appetite for a solution to their data privacy concerns, and (3) that Databox is the solution they want. Trialling the technology could help provide some answers to both questions. First, it would help to demonstrate



	the data storage and processing ability of the Databox, as well as the process by which companies would interface with it. Moreover, it would test the user interface and the use case, as well as user availability to consciously share personal data. The trial should be done with a sizeable number of paid volunteers covering different demographics.
Work with regulators	Liaise with regulators (e.g. the Information Commissioners Office) to understand how they would see the Databox contributing to the implementation the new GDPR. Focus on understanding the exact implications of the regulation, and to what extent the Databox can provide a solution for companies. Working with the regulator, try to find a compelling application of the Databox that make its value proposition clearer and more specific, steer the further development of the product, and better define the customer market. For instance, you can focus on sectors or organisations dealing with sensitive data (e.g. health, personal finance etc.) and/or sectors with high-value transactions (e.g. insurance industry, financial products etc).

5.3 Conclusions

5.3.1 Value proposition for companies

The analysis validated some of the hypotheses discussed at the beginning of this study. Both data-rich and data-poor companies are worried about the GDPR and its financial and/or reputational implications. However, uncertainties remain over the precise impact of the GDPR and what the best approaches to achieving compliance are. In this context, more can be done to articulate the Databox's contribution to achieving compliance, as well as the direct (e.g. technical, legal) and indirect (e.g. reputational) benefit this would generate for companies.

The main value of the Databox for the companies we considered is likely to be its ability to provide better insights into customer behaviour (especially for data-poor companies) rather than legal or financial protection. Companies would be happy to have access to insights as opposed to raw data but they would need strong assurances that the insights provided are accurate (by being able to verify the identity of end users and by having access to the data processing standards used), flexible (insights into diverse issues/questions) and easily accessible (through standardised, efficient processes). However, interviewees voiced concerns that data-rich companies could see the Databox as a threat to their business model or competitive advantage, and could in turn try to undermine it.

Interviewees suggested that companies are not prepared to pay for the Databox at this stage, but that they would value the insights themselves. Companies selling products that store and process personal data in the household could potentially work as partners in testing and distributing the Databox (e.g. by integrating it into one of their products) if they see strong customer demand or financially measurable benefits in achieving regulatory compliance.



5.3.2 Value proposition for individuals

Interviewees indicated that the demand for the Databox must come from individual consumers first, but that most **consumers have not yet shown sufficient awareness** of the problems concerning data privacy, or a willingness to pay for a solution to this problem. There is a risk that the Databox is providing a solution for a problem customers do not (yet) know they have.

There was agreement that creating a **B2C** data market could be an effective strategy to create value for individual databox users, but interviewees questioned the financial viability of this market given the cheap cost of data under the current model.

Interviewees also stressed the importance of creating an **intuitive user interface** that minimises decision-making by the users and simplifies the user experience (e.g. by using smartphones). However, it was felt that there is potential to create an intuitive user interface and that the focus of the Databox team at this stage is to validate the value proposition.

In summary, the value proposition and commercial viability of the Databox remain unproven, and it would represent a high risk play for any potential investor. The Databox does not seem to be a compelling proposition for the companies we have spoken to, as it is perceived to be ahead of the innovation curve and thus **preceding demand from the market**. However, there is openness to the value of different applications of the technology – especially as a means to provide new and better insights into consumer behaviour (e.g. polling machine model). **Speaking directly with consumers** would provide a better understanding of the ways in which they could engage with the Databox, whether it is solving a problem they recognise, and what direction the product development should take.

"We only need one more PR disaster for people to understand how compromised their personal information is and to start demanding better protection"