MANAGING DIGITAL TRANSFORMATION

Per Andersson, Staffan Movin,
Magnus Mähring, Robin Teigland,
and Karl Wennberg (eds.)

Managing Digital Transformation

Managing Digital Transformation

Per Andersson, Staffan Movin, Magnus Mähring,
Robin Teigland, and Karl Wennberg (eds.)

Karyn McGettigan, Language Editor

SSE INSTITUTE FOR RESEARCH

Stockholm School of Economics Institute for Research (SIR) is an independent research foundation established in 2010. The Institute's purpose is to conduct high quality academic research in the economic sciences with an ambition to combine scientific rigor with empirical relevance. Its board includes professors and other representatives of the faculty at the Stockholm School of Economics. The Institute encourages and assists its researchers to communicate their research findings to both the scientific community and society at large.

Address:
Stockholm School of Economics Institute for Research (SIR)
Box 6501, SE-113 83 Stockholm, Sweden
Visiting address: Sveavägen 65, Stockholm City
Phone: +46(0)8-736 90 00
www.hhs.se/en/Research/Institutes/institute-for-research/
publications@hhs.se

Keywords: digital innovation, organizational transformation, digitalization trends, customers, business models, platforms, eco-systems, analytics, information technology, change management, Internet of Things

Managing Digital Transformation
ISBN: 978-91-86797-31-7

First edition

Art Direction and Design: Petra Lundin, Manifesto
Production: Manifesto, www.manifesto.se
Cover photo: Westend61/Getty Images

Distributed by:
Stockholm School of Economics Institute for Research (SIR)

Printed by: BrandFactory, Göteborg, 2018

PROJECT SPONSOR

Stiftelsen Marknadstekniskt centrum, MTC
The Foundation MTC promotes value-creating interaction and learning between business and research in the areas of market, service development, digitalization and ecosystem development. The foundation was established by the Royal Swedish Academy of Engineering Sciences (IVA) and the foundation of the Swedish Institute of Management (IFL) in 1974. MTC is a non-profit organization, thus the projects are financed primarily by major corporations and government agencies.

mtc
STIFTELSEN MARKNADSTEKNISKT CENTRUM

In his central role at the Wallenberg Foundations,
Peter Wallenberg Jr has furthered a broad range of important research and research-led education initiatives at the Stockholm School of Economics (SSE) and its Institute for Research (SIR). This indispensable work has also helped create a fertile ground for research on digital innovation and transformation: a phenomenon currently experienced, shaped, and managed in and between organisations and throughout society.

This is the topic of this book, which we dedicate to him.

Contents

Acknowledgements

Every year since 1992, the Stockholm School of Economics Institute for Research (SIR) has published an Annual Research Anthology, and this year SIR is publishing the book in cooperation with MTC (Stiftelsen Marknadstekniskt Centrum). The purpose of the SIR Annual Research publication is to enable managers and practitioners better understand and address strategically important challenges by showcasing SSE research on a selected topic of importance for both business and society.

This year's book, *Managing Digital Transformation,* features authors from academic areas across SSE together with representatives outside the institution. The book's eighteen chapters show the strength and breadth of SSE's research within the area of digitalization and reflect the importance that SSE places upon closely linking research to practice and on investigating the leadership challenges and their implications in order to support value creation in society.

Participating in the many ongoing research projects at SSE and the multitude of aspects of digital transformation addressed in the various chapters has been very rewarding for the editors. We would like to thank all the authors for their hard work and cooperation throughout the project. In finalising this book, we have relied upon the expert work of Karyn McGettigan for language editing, Petra Lundin for layout and graphic design, and Marie Wahlström for digital access to the book. We are, indeed, most grateful for their excellent and diligent work.

The Director of SIR, Johan Söderholm, and the Chair of SIR, Professor Richard Wahlund, have provided important support, for which we are deeply grateful. We would also like to thank Vinnova for its financial support for the research project *Progressive Digital Development: Pre-Requisites for Success* of which this book is part. Hopefully, the book will become a reference for future research and funding areas.

Finally, we would like to thank all the companies and organisations for sharing their challenges and engaging in dialogue and research collaborations with us so that we can produce more solid and relevant research to help better our society.

The authors and the editorial team would like to express their gratitude to the following for generously contributing to this valuable research:

• European Union's Horizon 2020 research and innovation programme: Grant Number 688670
• Forte: Grant Number 2014-1502
• Hakon Swenson Foundation
• IIS: The Internet Foundation in Sweden
• Infina Foundation
• IoT Sverige: Internet of Things Sweden
• Jan Wallander and Tom Hedelius Foundation
• Marianne and Marcus Wallenberg Foundation
• Peter Wallenberg foundation
• SMHI: Sveriges meteorologiska och hydrologiska institut,
• Sveriges Riksbank
• Tore Browaldh Foundation
• Torsten Söderberg Foundation
• Vinnova

And to the Swedish retailers, executives and other research participants, and all others who kindly helped make this book a reality.

Stockholm, January 2018
Per Andersson, Staffan Movin, Magnus Mähring, Robin Teigland, Karl Wennberg

Introduction

One of the hottest research topics lately is digitalization. Many research projects are focusing upon different perspectives. Gone are the days when digitalization or business implications of ICT were just about increasing efficiency. Instead, the ripple effect of digital development can now be felt wider and deeper than ever before. The way in which business is conducted and how it creates value, as well as how corporations can become more efficient and sustainable, are all implications of digitalization. Adapting to new demands and taking advantage of the plethora of possibilities, however, is not always easy.

Managing digitalization and the transformation of business always involves new challenges. The novelty and complexity of the digital age has led to an increased academic interest in the area of digital transformation and a call from companies that seek support in this process.

We take a look at digitalization from the perspective of business research. This creates a better understanding of the challenges that today's businesses are facing. We believe this anthology will serve as a tool to help businesses better understand the force that is digitalization and support these corporations in their digital transformation.

The idea behind this anthology grew as Marknadstekniskt Centrum was taking part in several interesting research projects. Companies were asking MTC to facilitate contact with scholars and supply them with academic insight. Vinnova came on board, by supporting the project *Progressiv digital utveckling förutsättningar för framgång (Progressive Digital Development: Pre-Requisites for Success)* of which this book is a part: its aim to stimulate business to become more progressive in digital change. At last, this book and the website www.digitalchange.com have become a reality.

This joint venture between Marknadstekniskt Centrum and The Stockholm School of Economics Institute for Research follows the SIR tradition of publishing an annual yearbook to showcase its vital research contributions. The book begins with an overview of digitalization, then moves to understanding the new digital customer, and ends by exploring re-organisational effects, business models, and ecosystems. We hope this year's anthology will be useful for managers by facilitating their digitalization processes.

PART 1: DIGITALIZATION – DIFFERENT PERSPECTIVES

The role of digital technology in business and society is rapidly shifting from being a driver of marginal efficiency to an enabler of fundamental innovation and disruption in many industrial sectors, such as media, information and communication industries, and many more. The economic, societal, and business implications of digitalization are contested and raise serious questions about the wider impact of digital transformation. Digitalization affects all private and public operations, as well as the internal and external workings of any operation. Digitalization is the major driving force behind sweeping large-scale transformations in a multitude of industries. Part 1 includes various perspectives on digitalization and digital transformation.

PART 2: THE NEW DIGITAL CUSTOMER

Digitalization has resulted in more user-centric business and user-centric systems. The changing behaviour of the digital consumer/customer is discussed here as it connects to new forms of customer involvement and engagement, as well as analysis models of what creates customer value in this digital context.

PART 3: THE RE-ORGANISATION IN ORDER TO CONNECT WITH THE DIGITAL CUSTOMER

How can companies connect with digitalized consumers and non-digitalized customers? This is a central issue in managing digital transformation, as it draws attention to the emerging intra-organisational, marketing, and customer interaction challenges associated with digitalization: for both the consumer and the supplier. Another aspect of this is the internal handling of new forms of organizational ambidexterity; that is to say, companies and organizations engaged in digitalization processes often require an internal re-organisation in order to handle the demands that digitalization brings, and to explore new digital opportunities while promoting their existing business and operations.

PART 4: BUSINESS MODELS AND ECOSYSTEMS

How do companies change, adapt, and innovate their business models? Given that digitalization leads to a convergence of previously unconnected or loosely connected markets, the digitalizing company and organisation is analysed in its systemic and dynamic context. This part draws attention to business models

and business model innovation. Incumbent firms need to adapt and change business models while competing with digital start-ups based upon new scalable business models, accessible ventures, and rapid processes of intermediating. These chapters discuss completely new co-operative business models: processes that need to be developed as companies shift from products to digitally based services.

The Ecosystem places digitalizing organisations and companies into their broader and systemic context. This includes discussions on digital disruption, industrial convergence processes, and shifting patterns of competition and cooperation. Digital technologies cause markets to converge in many new and sometimes unexpected ways. The result is the emergence of new roles and market positions of technical platforms.

Staffan Movin, Stiftelsen Marknadstekniskt Centrum

Digital Trace Data: Which Data Should we Collect and What Should we do Once we Have it?

CLAIRE INGRAM BOGUSZ

Introduction

Today, we take for granted the fact that websites and other platforms are optimised to provide different experiences for different people. Individuals' browsing history, location, and device are all bits of important information that help web services provide them with a personalised experience. These personalised experiences are incredibly useful for individuals; someone on a mobile phone no longer has to navigate a complex web-based version of a platform, for instance, because data about the individual's device are received by the website provider. Similarly, advertisements that individuals see are tailored for them, based upon what they have looked at – and read – online. While there is no question that these developments are convenient, some question whether data collection has gone too far.

Whether or not we are aware of it, these practices form the very backbone of some of the largest internet firms. Google, for instance, makes much of its revenue from advertising: in mid-2016, Google's parent company Alphabet made 21.5 billion USD in revenue, of which 89 per cent (or 19.1bn USD) came from advertising (Johnson 2016). Facebook made 8.81 billion USD profit in 2016, exclusively from advertising – with 84 percent of that coming from mobile advertising (Constine, 2017).

While there are ways to avoid generating data while online, most of us do not take the trouble to cover our footsteps. This is despite the fact that everything from how our mouse moves when interacting with a website, to sites

that we visit, to the data that we enter into online forms – even if we never submit the form – can be, and often are, collected.

The data that enable personalisation are generated on a diverse number of websites: from social media to news. And, all of these websites collect these data, whether it is for their own uses or otherwise. When it comes to their own uses, many websites pay for their own existence by selling advertising, and they use our data to match our profiles with the most relevant advertisements. In principle, this should give us the most relevant advertisements online; however, the process also allows websites to charge advertisers more for better targeting. Other actors who make use of these data are third party actors; they bundle data from multiple sources and sell them to other firms: sometimes in raw form, and sometimes as analytical insights. This direct and indirect collection and use of online data has come to be called the *commodification of data*, and it has emerged from our desire to have "free" services on the web, and the fact that digital footprints – or digital traces – are easy to track.

This chapter will define and give boundaries to the collection of what is called *digital trace data*: the data that are generated – and collected – as a by-product of our online activities. This summary then goes on to show how digital trace data are being used for both business and illicit purposes, and zooms in on some examples. Lastly, it discusses how to balance the risks for individual integrity against the opportunities for new businesses, and "free" services on the web.

What are "Trace Data"?

"Big Data" is the popular term for very large volumes of data; it is often used to refer to the volumes of data collected about online activities. However, big data has its origins in scientific inquiry, and internet-generated data is just the tip of the iceberg: information collected about the weather patterns by satellites, to interaction data from so-called "Internet of Things" devices connected to one another, to information about online activities are even more commonplace (McAfee et al., 2012). These large volumes of data have led to more accurate and sophisticated models in areas such as agriculture and weather patterns (using weather data), allowed connected devices to make predictions about when to turn the heating on in our homes, or when to refill the refrigerator (from Internet of Things data), to complex models around human behaviour and preferences.

Large volumes of data allow for automated pattern recognition, the testing of hypotheses at a rate of knots, and automated model development using machine learning. When it comes to online activities, the more gritty "digital trace data" is the most relevant for social – and economic – activities. Individuals typically unintentionally leave digital traces as they browse, shop, and transact online. These data can be used instead of, or as a supplement to, data already willingly given by individuals in order to build a clearer picture of their online activities and preferences.

Most of us are aware, at least peripherally, that we generate data while online. Most websites display, for instance, a "cookie request": they ask for permission to store small amounts of data on our computers. These small files are linked to a particular website; in turn, the files can be accessed both from the user's computer and from the website owner's server. As such, the files carry information that is used to fine-tune the user's online experience by remembering preferences or providing targeted advertising. Often web pages contain scripts that allow data to be carried from one visit (or page) to the next: for instance, to optimise advertising.

Cookies are just the tip of the iceberg: not only are we mostly aware of them; there are limits on what can be shared and are regulated by bodies such as the European Union (Directive 2002/58/EC). Other traces left online are not as tightly controlled. However, to understand why (and how) this is the case, we need to explore what it is that we are talking about when discussing digital traces.

These different ways in which data are collected, the different kinds of data, and the extent to which we have control over the data collected allow us to classify digital traces to some degree. Drawing upon Schneier (2015) and Ingram Bogusz (2018), this chapter describes a taxonomy of digital trace data, and discusses the different ways in which these data are being used commercially, both on their own and in combination with other data. This summary then goes on to discuss the possibilities and pitfalls of data use.

DATA WITH CONTENT, AND DATA ABOUT DATA

Data typically are one of two kinds: data with content or metadata. Data with content are substantial and personal in nature; they say something about an individual, and can easily be identified as being linked to that person. Data with content include not only that which we explicitly share with a firm; they

also include other kinds of trace online data, notably content shared on social media and in forums. When it comes to social media, for instance, photographs on Facebook are data with content, as are the links to news articles that we share.

In contrast, metadata are data about data. For instance, metadata around a Facebook photo might include the size of the photograph, the time it was shared, and the IP address from which the image was shared. While the metadata from a single photograph cannot be used to say much about an individual, the metadata about all of the photographs of you shared on Facebook can. For instance, if you consistently share large files, from the same IP address in Stockholm at the same time on a Friday night, algorithms might determine that you have a high quality camera and therefore are a photography enthusiast, who lives in Stockholm, and prefers not to go club-hopping.

Many companies – and countries – treat the collection of metadata (and other "anonymised" data) as unproblematic. Indeed, metadata are often central to, for instance, a telecommunications firm, which ensures their internet service infrastructure is working as it should. However, depending upon the patterns searched for in the data, it could reveal more about individuals than data with content. Therefore, these data could give firms, and anyone else who can access these data, an unprecedentedly detailed picture of a number of online habits. For instance, the presence of a mobile phone at an anti-government protest in an autocratic country might reveal the identities of individual protestors.

A recent study, for instance, using only the metadata from phone calls and text messages identified that a small sample of individuals were suffering from sensitive medical conditions (Mayer et al, 2016). The amount of data that is currently available about us, combined with advances in data analysis, have significantly increased the likelihood that an individual can be re-identified from anonymised data: whether metadata or otherwise. In fact, removing personal data from digital traces (for instance, by making it illegal to collect personal data) is, therefore, insufficient: identifying an individual depends upon the number of data traces available, and that to which other data a dataset can be linked.

LEAVING TRACES IS ALMOST UNAVOIDABLE

Most of us are familiar with giving some of our data to companies and authorities; it would be impossible to get a bank account, or access healthcare, without disclosing our names, addresses, and other information. We are fully aware of these data, however, and what they say about us. This is not always the case with digital trace data. Although the comparison is imperfect, digital trace data has been likened to the data left behind at a crime scene.

Take, for instance, when a perpetrator leaves behind a strand of hair or fingerprints on a doorknob. Criminals are not the only ones that leave these traces, and neither criminals nor passers-by leave these traces deliberately. Even so, they can be used to identify individuals–whether at face value by their appearance or colour, or after further analysis. Thus, these traces may be said to have "content".

Avoiding leaving these traces in the physical world, while possible, is tricky. A perpetrator could wear gloves to avoid leaving fingerprints, or a hat to avoid leaving hair, but what about footprints or skin cells? It would take considerable effort, if possible, to avoid leaving any traces at all. Leaving these traces in the digital realm is similarly difficult to avoid: online activity can be re-routed through multiple servers, and users can use special web browsers. Few, however, go to these lengths.

For this very reason, online data are readily accessible. While we give explicit permission for some services to make use of our data – for instance, Google and Facebook – data are accessible even to those to whom we do not give explicit permission. We also often give implicit permission for our data to be used and stored, simply by using certain websites: such as forums or news websites.

From Where do They Come?

These two overarching categories are the ones most often discussed in policy documents, for example. However, individual users have different levels of control over data, depending upon who shares them and who controls what is being shared. For this reason, researchers have further characterised data. While there are many taxonomies, we favour Schneier's (2015) in this chapter. Relying upon his framework, we discuss some of the different types of data out there, their source, and who controls the data in various cases.

SERVICE DATA

We are largely familiar with giving our service data to service providers: one's name, age, address or credit card number are common service data. Indeed, these data are willingly given in the offline world to everyone from banks to state actors. However, despite how widely these data are used, they are considered to be very sensitive. Ironically, although we often willingly provide these data, they are the most heavily protected in most countries; for instance in Sweden, other data may not be used to infer these personal details.

These service data, in the digital world, have the least use: they give only the most basic of details about an individual, and information contained in service data can often be inferred from other data. For instance, someone's location could as easily be inferred from an IP address as his or her physical address. Moreover, an IP address can pinpoint where an individual is located at a given time, and not just what their home address is: useful information if, for instance, for a targeted advertisement for coffee at 7am.

Considering how service data has been the backbone of the service industry for decades is also illuminating. Today, other sources of data are more enlightening than this service data. Therefore, it is interesting that individuals (and governments) are protective of these data, when other kinds of data contain the same information: often in real time. Other sources of data often reveal more about an individual than a name and address can.

Theses service data, however, can be combined with other sources of data, and others' service data, to create data that provides more – and deeper – insight.

DERIVED DATA

Derived Data are data inferred from other data. For instance, combining service data from thousands or hundreds of thousands of individuals allows marketers to create segmentations. Offline brokering firms create group profiles that categorise people according to their shared demographic traits, while online information brokers tend to use social media networks, device locations, and online activity.

The creation of these categories is done by individuals or machine learning, with no input from the individuals whose data are being curated. Thus, an individual's membership of a group created based upon either demographics or online activity is not something that they can influence: even if that

categorisation is inaccurate or just plain wrong. Moreover, as third parties create these categories, the individual user has no influence over how these group-level data are used by the brokering company or other third parties.

DISCLOSED DATA

Disclosed data, on the other hand, are data that include content that we as individuals control, according to where we control the platform. This kind of data includes content such as photos, messages, and/or comments that we post on a webpage, blog, and/or website that we control, own, and/or host. While the data are publicly available, we can decide what to share, and for how long. In principle, this should mean that we could limit access to the underlying infrastructure, thus, limiting the collecting of digital trace data by third parties. However, the reality is the "public" nature of these data – even though we control the content – mean that third parties wanting to use it can easily do so as well.

We often think of data that we put up on social media sites as being disclosed and within our control. However, this is not the case. Even data that we flag as "private" can be used by social media giants for third party services, such as advertising targeting. For instance, up until 2016, there was talk of using Facebook data for credit scoring[1].

ENTRUSTED DATA

What we often think of as disclosed is really entrusted data instead. This includes similar content to disclosed data, yet it is data posted on a platform we do not control, such as Facebook, LinkedIn, or our employer's website. As such, someone else decides what happens to these data, and how easy they are to use and collect. We can decide the content and whether or not we chose to post it on these platforms; however, we cannot control what firms subsequently do with our trace data.

Entrusted data has been a goldmine for internet giants. For instance, by making use of entrusted data, Facebook has built some of the world's most reliable facial recognition software. By using photographs online and users' tags of their own friends, Facebook has been able to teach a machine-learning algorithm how to recognise and classify facial features. This algorithm is now

1 http://fortune.com/2016/02/24/facebook-credit-score/

not only better than humans are at facial recognition; its use online has been called "biometric invasion of privacy" in court proceedings that aim to curb its use (Brandom, 2016).

Facebook and other social networks are also renowned for their use of data generated as a result of this entrusted data, namely incidental data.

INCIDENTAL DATA

Incidental Data are data generated as a result of the sharing of entrusted data. For instance, comments on photographs or on shared links are incidental data. The tag on a Facebook photo, which identifies an individual, is also considered incidental data. Incidental data, as with entrusted data, is beyond the user's control: both because of its platform and because it is generated by a third party.

Incidental data in the business world are often used to train machine-learning algorithms or to generate business insights. The example of Facebook's tags is an instance of algorithm training. This data can also be used to generate business insights when analysts use natural language processing to assess whether or not a post or online content has been positively or negatively received – not just whether or not it has been shared.

BEHAVIOURAL DATA

Lastly, there are behavioural data. These data are created while interacting with a computer, mobile phone or tablet. Some examples include how long one spends looking at a particular website or where one clicks. These kinds of data provide insight into what we do, with whom, how often, and where. These behavioural data are among some of the most valuable data to collect: they allow websites to give individuals tailor-made advertisements or special offers.

Behavioural data have even been used to conduct credit risk assessments. Wonga, a payday lender in the United Kingdom, claims that its behavioural data-driven algorithms are so reliable (and quick) that decisions are made within six minutes, and that money is transferred to user accounts in fifteen (Deville, 2013). Wonga does this by tracking how a person uses a sliding credit bar (dragging it straight to the maximum amount is apparently a red flag). Moreover, Wonga seems to offer individuals higher initial loan amounts based upon the device from which they access the site.

Table 5.1: The Characteristics and Kinds of Digital Trace Data, from Ingram Bogusz (2018)

	Deliberately Left	Unintentionally Left	Left by a Third Party
Data with content	Service data Disclosed data Entrusted data		Entrusted data Incidental data
Metadata		Entrusted data Behavioural data Derived data	Incidental data Derived data

Both behavioural and incidental data are typically unintentionally (or unknowingly) shared. This commonly occurs when we allow one service access to data contained in other services; for instance, when we allow the Facebook mobile app to access our phonebook, we ultimately are sharing our friends' phone numbers. The fact that data are unintentionally shared, however, does not affect who has control over when and to whom, data are released.

All of these kinds of data could be either data with content, or metadata: that is to say, data about data. We unknowingly generate this metadata over the long term in an organised format.

Having discussed the volumes of data that we generate, and the differences between them, we now turn to the broader trend of commodifying data, before discussing how to approach the possibilities implicit in these data with privacy in mind.

Patterns in Commercial Trace Data Use

Facebook's average revenue per user in the US and Canada was around $20 in 2016 (Oreskovic, 2017). This revenue is largely a result of the social media giant's access to volumes of data that, at scale, it can use to create insights and new products. These new products include targeted advertisements, news feeds that contain "recommended" posts and new software, such as the aforementioned facial recognition software.

What is key is that individual data sets are not worth this money: data can only be used to build new products and train algorithms when an actor controls and maintains vast quantities. While building and maintaining this infrastructure costs money too, the value of the data is only growing: the European Commission (2016) estimates that, by 2020, the value of European

citizens' personal data is expected to reach 1 trillion EUR, or 8 per cent of the Union's GDP.[2]

DIGITAL INSIGHTS AND RECOMMENDATIONS

Facebook, Google, and other social media giants are, at their very core, data brokers. They use entrusted and incidental data to build profiles of individuals for various purposes. They also access other data to know where we are, such as our phones' GPS position. These profiles and physical world indicators allow them to create some of the following business and service innovations.

Consumer Segments

Segmentation helps retailers online and offline identify potential customers: for instance, "under 40 without a mortgage" or "young mothers in the Uppsala region". These profiles can then be sold to other companies for advertising or marketing – whether through the data broker's platform or otherwise. The media giants with whom we are familiar, however, are just the tip of the iceberg. There are even more data brokers of whom we have not heard: US–based company ID Analytics has information on more than 1.4 billion consumer transactions. The data to which they have access goes far beyond what Facebook or Google control; instead, they can offer third parties detailed pictures of consumer browsing and purchasing practises, their interests, habits, hobbies, communities and opinions.[3]

Consumer Behaviours

Knowing whether a visitor to a site is a "first time visitor" or "everyday browser (who never buys)" can be vital information for an online retailer. These categories help online platforms optimise their appearance – and offerings – for different people, depending upon their internet profile and browsing history. This information, and the resulting personalisation of platforms, is both useful for individuals and for retailers: it allows the platform to better meet the consumer's needs that, in turn, increases its own income.

2 http://ec.europa.eu/justice/data-protection/files/data-protection-big-data_factsheet_web_en.pdf

3 Federal Trade Commission (2014) 'Data Brokers: A Call for Transparency and Accountability.' Washington DC: Federal Trade Commission. p.iv.

NEW PRODUCT DEVELOPMENT

Training Algorithms

Algorithms using machine learning are increasingly common online and offline. Google and Uber's self-driving cars, for instance, are steered by algorithms that have been "trained" using environmental feedback. Other kinds of algorithms are often designed – and trained – using online data: for instance, ones that automatically calculate credit risk, set prices or recommend products.

These algorithms are given initial instructions, often in the form of experiments, which rely upon one set of data to complete. Based upon initial data and instructions, a machine-learning algorithm builds a model of some sort. This model is then tested using either additional data ("supervised modelling") or user feedback (usually "unsupervised modelling"). Therefore, being able to build these kinds of algorithms requires access to large volumes of data. And the larger the volumes, the more accurate the algorithm is likely to be. In a study by US credit assessor FICO, using machine learning was said to improve the accuracy of a credit assessment by 10-25 per cent, depending upon the methods used.[4] One caveat to this is that only the right kind of data can generate these results: not only are some data not inherently useful; the cost of extracting them may be more than the possible benefit they reap.

Making Processes into Products

Building a credit score today is something that is based upon transaction and financial activities, as well as service data. Individuals build up credit scores by borrowing and repaying progressively larger amounts of money: by consistently having their salaries paid into a single account, and by paying bills on time. Financial institutions can lend a consumer money at a given interest rate, based upon these and other consumer-disclosed or service data: with higher interest rates correlating to high-risk lending. However, these data points provide only the broadest frame for assessing an individual's creditworthiness.

Moreover, behavioural and incidental data can provide a clearer picture of an individual's disposition to repay a more accurate credit score. Third parties

4 http://www.fico.com/en/blogs/analytics-optimization/how-to-build-credit-risk-models-using-ai-and-machine-learning/

– including data brokers – have, therefore, turned the calculation of these scores into new products. The ability to screen potential borrowers more accurately and possibly more quickly than competitors is, indeed, a source of competitive advantage. The provision of these kinds of products can draw new potential customers into the credit ecosystem. In China, for instance, the use of digital traces has meant that people who were once ineligible began to get credit, which served to the benefit of the economy at large (Bateman 2017).

Discrimination and Profiling

While creating consumer profiles for the purposes of providing personalised services may seem sensible, these methods can sometimes be used to "profile" individuals for nefarious purposes. As a test, a research team at Stanford University recently created an algorithm that, by using public images of faces, could identify the sexual preference of the person in the image. Moreover, the algorithm was more accurate than the average human.[5] The abundance of public data has meant that, while algorithms can be created to do commercially and socially valuable things – such as track weather patterns and identify health risks–the data that are out there are use-agnostic. Therefore, it is possible for data to be used to support immoral and even dangerous developments.

An algorithm that identifies homosexual men or women is just the tip of the iceberg. Given how digital traces are increasingly used to train machine-learning algorithms, even the creators of algorithms lose control of what it is their creations do with the data – and what kind of heuristics they create. Algorithms used in hiring decisions have been observed to adopt human biases because the data upon which they rely contains these biases.[6] One price-setting algorithm has been known to use race as a proxy for academic achievement.[7]

What is more, these algorithms often self-teach ("unsupervised modelling"), making it unlikely that a human would notice and figure out how to reverse discrimination.

5 https://www.wired.com/story/ai-research-is-in-desperate-need-of-an-ethical-watchdog/

6 https://hbr.org/2016/12/hiring-algorithms-are-not-neutral

7 https://www.propublica.org/article/asians-nearly-twice-as-likely-to-get-higher-price-from-princeton-review

Implications for Businesses

Most businesses today have a digital presence of some sort – whether through their website, digital advertising, or online customer service (often all of the above). This means that most businesses have the potential to benefit from the data that is being generated, which is close to their brands, products, and services. A few things to think about include the following:

OBTAINING THE DATA MAY COST MORE THAN THEIR WORTH

Data from existing products and services can be hard to access, largely because online services built in the past did not prioritise data in the same way. Thus, data are often not collected; they may be siloed or only a very limited dataset collected. Business owners may be misled into thinking that all data are gold; however, this is not true. Only data that are accessible, and relevant to a given question, can improve business outcomes.

BUILD A NEW SERVICE OR PRODUCT WITH DATA IN MIND

Although there are vast troves of data out there, accessing them can be messy and expensive. For this reason, businesses building new digital services are advised to construct them a little more slowly, and think about the kind of data that might be collected – and the way in which they might be used as part of the service or product innovation process. Collecting and structuring the data in a useful way right from the beginning makes it easier to work with later; this maximises its potential.

INFORM YOUR CUSTOMERS

Not only does informing your customers that you are collecting their data build trust, it is currently required by a myriad of laws. However, telling them what you are using it for is not yet required. This will shortly become mandatory as a result of the European General Data Protection Regulation (GDPR), which comes into force in Sweden in 2018. However, even if it were not required, it makes business sense to be transparent about the way in which you use your customers' data: in order to avoid a backlash.

DATA PROTECTION BY DESIGN

The GDPR also encourages what consumers are increasingly demanding anyway: what is known as "data protection by design". This principle encourages

architects of data services to use techniques such as anonymisation, pseudonymisation, encryption, and other protocols for anonymous communications. The European Commission encourages the use of these techniques as *ex ante* protection from data violations, and has offered to support member states in the technical implementation of such measures.

Moreover, these laws have also tried to avoid lax data security by making more severe penalties for data breaches; the GDPR prescribes a fine of up to 20 million EUR or 4 percent of a firm's global turnover (whichever is higher) for companies who misuse personal data or fail to take proportionate steps to prevent data breaches. This protection also helps to deal with some of the problems previously identified around metadata.

Policy Considerations

Although the abundance of digital trace data allows for the creation and optimisation of large numbers of new services, using these data runs the risk of infringing on individual integrity. In examining digital trace data and legislating around its collection, storage, and use, regulators must find a balance between the commercial imperative to support new business creation and the social necessity to support individuals' data integrity. This section is devoted to some of the important elements that data legislation should – and increasingly does – include for the purposes of individuals' protection.

INFORMED CONSENT

While consumers are often told that their data are being collected, it is not always clear for what purpose – or exactly for which data – they are giving their consent. In fact, the use of machine learning may even mean that corporate data scientists are not always sure themselves what it is that their algorithms are prioritising.

In 2008, researchers at Stanford University estimated that it would take 7.6 days per year for the average person to fully read all the privacy statements they encountered in their lives (Symons & Bass, 2017). In practice, the "take it or leave it" of most online service terms and conditions means that people have no choice but to grant access to their data to a large number of compa-

nies. That is to say, there is no option to make use of, for instance, Google's services without permitting them to analyse and sell the data they collect.[8]

Moreover, most of us barely acknowledge (and seldom read) notices around how our data are collected online. In fact, the norm when using a website is often just to "accept" the terms and conditions of its use, without reading what they entail. This means that users often do not know that their digital traces are being collected, and do not know what are the ways in which the data are being used nor to whom they might be sold.

Users typically make use of services without being able to limit the extent to which data are collected and used, even if they were aware of it, which many are not. Moreover, even if individuals were aware that their data was being used, they would be hard-pressed to understand how its use would affect the financial (and other) services they receive.

The GDPR requires that individuals give specific and informed consent to how their data are used and collected. While this is a move in the right direction, the complexities of algorithms and the fact that individuals seldom read online terms of service means there is still more to be done.

While it is hard to legislate or avoid over-use of user data when consent is given, GDPR has also promoted better safety measures to prevent the theft of data, and the non-consensual identification of individuals.

CREATION OF "DIGITAL COMMONS"

While the GDPR goes a long way toward protecting individual data, it has been suggested that individuals should have control over digital traces about themselves online: for instance, in the European Court of Justice's 2014 ruling on the "Right to be Forgotten".[9] In the UK, one proposal has been to create a registry of data used by firms (Downey, 2016). In Australia, draft legislation has proposed a National Data Custodian body to allow individuals to have greater control over the data collected about them by both public and private sector actors (Bindi 2016). Germany, known for the importance it places on privacy, treats data protection as a consumer protection issue, with breach offenses under the law.

8 Some uses of the data can be limited, but users seldom know that these limitations exist—or how to make use of them

9 ECLI:EU:C:2014:317

Pentland (2013) suggests that our digital trace data should be managed by data controllers in a way akin to how our banks manage our money. He highlights the tenets of possession, use and disposal, arguing that these are the three areas of digital trace data leverage that should be regulated and overseen. He describes these tenets as follows:

> *You have the right to possess data about you. Regardless of what entity collects the data, the data belong to you, and you can access the data at any time. Data collectors, thus, play a role akin to a bank, managing the data on behalf of their "customers.*
>
> *You have the right to full control over the use of your data. The terms of use must be opt-in and clearly explained in plain language. If you are not happy with the way a company uses your data, you can remove the data–just as you would close your account with a bank that is not providing satisfactory service.*
>
> *You have the right to dispose of or distribute your data. You have the option to have data about you destroyed or redeployed elsewhere. (2013:37)*

An experimental project called DECODE (DEcentralised Citizen-owned Data Ecosystems) is currently underway with partners in Spain and the Netherlands; it aims to develop technology to facilitate this "data commons". The intention is to put people in control of their personal data, and give them the ability to decide how it is shared. The technology will include an architecture for controlled and, if desired, anonymised data sharing. Crucially, this project also explores whether there are viable alternative revenue generation models in an internet economy, which finances itself predominantly through monetising personal data

Conclusion

As more economic activity has moved online, records of our activities have improved and are routinely collected: both with and without our consent. Much of the activity that precedes an economic transaction is also recorded: individuals' attitudes to brands, online habits, and even decision-making processes can be captured through the things they say and do online. This personalisation of online experiences extends beyond just how and what we see online, to what we are offered, by whom, and under which terms.

This chapter has explored the kinds of data that have been made available as a result of improved – and more determined – data collection by internet giants and data brokers. However, while this abundance of data has made

way for new products, services and analytics, it raises concern around individual integrity. While the EU's GDPR goes a long way toward easing these concerns, a push in the direction of a "data commons" would give consumers more control over their data. This is important and, thus, should be discussed in the long term.

References

Brandom, R. (2016, May 5). Lawsuit challenging Facebook's facial recognition system moves forward. *The Verge*. Retrieved from https://www.theverge.com/2016/5/5/11605068/facebook-photo-tagging-lawsuit-biometric-privacy

Constine, J. (2017, February 1). Facebook beats in Q4 with $8.81B revenue, slower growth to 1.86B users. *Techcrunch*. Available online at https://techcrunch.com/2017/02/01/facebook-q4-2016-earnings/

Deville, J. (2012). Regenerating Market Attachments. *Journal of Cultural Economy, 5*(4), 423–439.

Downey, P. (2016). Registers in a digital ecosystem. Available online at https://data.blog.gov.uk/2016/09/12/registers-in-a-digital-ecosystem/

Ingram Bogusz, C. (2018). Digital Traces, Ethics and Insight: Data-Driven Services in Fintech in Teigland,R; Siri, S; Larsson, A; Moreno Puertas, A; and Claire Ingram Bogusz (Eds), *The Rise and Development of Fintech: Accounts of Disruption from Sweden and Beyond* (forthcoming). London: Routledge.

Johnson, L. (2016, July 22). Google's Ad Revenue Hits $19 Billion, Even as Mobile Continues to Pose Challenges. *Adweek*. Available online at: http://www.adweek.com/digital/googles-ad-revenue-hits-19-billion-even-mobile-continues-pose-challenges-172722/

Lorenzetti, L. (2016, February 24). Lenders Are Dropping Plans to Judge You by Your Facebook Friends. *Fortune*. Available online at http://fortune.com/2016/02/24/facebook-credit-score/

McAfee, A., Brynjolfsson, E., Davenport, T.H., Patil, D.J. and Barton, D. (2012). Big data: The management revolution. *Harvard Business Review, 90*(10): 61–67.

Oreskovic, A. (2017, February 20). Facebook generated almost $20 from each of its US and Canadian users last quarter. *Business Insider*. Retrieved from http://nordic.businessinsider.com/facebook-almost-20-per-user-us-and-canada-q4-2017-2?r=US&IR=T

Pentland, A. (2013, October). Data-driven society. *Scientific American*.

Schneier, B. (2015). *Data and Goliath: The hidden battles to collect your data and control your world*. New York: WW Norton & Company.

Symons, T., & Bass, T. (2017). *Me, my data and I : The future of the personal data economy*. Retrieved from https://decodeproject.eu/file/158/download

Case C-131/12, Google Spain SL, Google Inc. v. Agencia Española de Protección de Datos (AEDP), Mario Costeja González, 2014 ECLI:EU:C:2014:317, 100(3) (May 13, 2014).

Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector. *Official Journal L 201, 31/07/2002 P. 37 – 0047*.