The Data Value Chain

June 2018

The GSMA represents the interests of mobile operators worldwide, uniting nearly 800 operators with more than 250 companies in the broader mobile ecosystem, including handset and device makers, software companies, equipment providers and internet companies, as well as organisations in adjacent industry sectors. The GSMA also produces industry-leading events such as Mobile World Congress, Mobile World Congress Shanghai and the Mobile 360 Series conferences.

For more information, please visit the GSMA corporate website at **www.gsma.com**

Follow the GSMA on Twitter: **@GSMA** and **@GSMAPolicy**

For more information or questions on this report, please email **publicpolicy@gsma.com**

A.T. Kearney is a leading global management consulting firm with offices in more than 40 countries. Since 1926, we have been trusted advisors to the world's foremost organizations. A.T. Kearney is a partner-owned firm, committed to helping clients achieve immediate impact and growing advantage on their most mission-critical issues.

For more information, visit **www.atkearney.com**

Contents

Executive summary

The availability of data is transforming the global economy, as well as large sections of civic society on an unprecedented scale. Many organisations, regardless of industry focus, now consider data to be a vital strategic asset and a central source of innovation and economic growth. A recent title page of The Economist[1] asked if data is the "new oil", in other words a raw material of great value in powering the economy. Instead, data seems to be a new form of asset, one where bringing together individual pieces increases the value by delivering new insights and correlations. In addition, using data does not consume that data: rather it can still be used by others, or built on and reused many times over. As a result, being able to collect and exploit large volumes of data in an efficient manner has become an important source of value. Data-driven businesses currently generate higher returns than traditional businesses since, once collected, the data can be mined for multiple purposes and cycles of collection and exploitation become self-sustaining.

1. The Economist, print edition, May 6th, 2017

The business of collecting and exploiting all types of data, whether personal, machine or system generated data, can be analysed with reference to a value chain framework. This emerging data value chain consists of several discrete steps:

- **Generation** – recording and capturing data;
- **Collection** – collecting data, validating and storing it;
- **Analytics** – processing and analysing the data to generate new insights and knowledge; and
- **Exchange** – putting the outputs to use, whether internally or by trading them.

In a traditional value chain, different companies would typically specialise in a limited set of activities and then trade inputs and outputs with other companies, with value created at each step of taking raw material inputs through to finished goods and services. However, the nature of data results in a tightly integrated value chain where the organisation that collects the data is very likely to retain it through the steps to develop the output themselves, while buying in specific supporting services. This report explores the reasons for this and the implications for value chain players and policymakers.

A key characteristic of data is the fact that in most cases it is hard to put a tangible value on a single piece of data as it advances through the value chain. Data has a high degree of private value uncertainty. The private value of data depends on the value of the insight that is being produced. As that insight might be unknown or the product of many data points, it can be hard to value a single piece of data. Any such trade would have high transaction costs. The buyer would need to verify the quality and validity of the data (and often obtain the permission of the data provider) so that it is much more efficient for a potential buyer to collect the data itself or via exclusive agents, rather than procure it from the open market. It is also true that for many uses it is necessary to build a very large dataset, often from multiple sources. It is only when mining and interrogating large datasets that the true value of the input or source data can be discovered, or even which pieces of data contributed to the outputs and which went unused. There are emerging examples that for specific applications these hurdles can be overcome as the value chain matures and trading practices become more established.

Due to the nature of data collection, many data-driven businesses benefit from significant scale and network effects. These can pass a tipping point and become self-reinforcing – which is sometimes referred to as 'winner takes all' conditions. The need to collect substantial amounts of data in order to be able to find the right subset or combination of data to commercialise can also lead to vertical integration that enables companies to expand the scope of data they collect. This is especially true of platform type services, which seek to bring together the largest communities of users (in the case of social media) or buyers and sellers (for transaction-based platforms). The fact that data can be put to multiple uses simultaneously, means that there are instances of companies that are able to build large datasets and then interrogate and leverage them to build powerful businesses in adjacent market segments. While such behaviour is a long-established business practice, the indirect network effects of the data value chain can give data-driven platforms a significant competitive advantage in largely unrelated services in ways that other players would find it difficult to replicate. For example, while search, e-mail and entertainment may seem unrelated services from a consumer perspective, being able to develop a profile of an individual's interests and shopping habits increases the value of such profiles to potential advertisers who can then co-ordinate how they target such an individual in a way not feasible via a standalone service.

This report acknowledges that integrated businesses and platforms are often the most practical and commercially efficient way – at least today – of providing such services and they yield genuine consumer benefits. Nonetheless several issues emerge that are very important when evaluating data-driven markets from a public policy and enforcement perspective.

1. **Direct and indirect network effects** and their impact on the potential for well-functioning markets need to be better understood, particularly the way some platforms collect and use data as part of operating their service but also potentially as means to entrench their position against competition. Social media services are a good illustration of the strong network effects, since a service with many users is likely to attract even more users and therefore generate revenues that can be used to improve, promote and scale the service, and thereby attract yet more users. It then becomes very hard for a competitor, regardless of how good or bad their service is in comparison, to compete directly. The success of Facebook in attracting over 2 billion monthly active users shows how far such network effects can extend via the global internet[2]. A recent European Commission fact-finding exercise[3] found that the strong platform effects in digital business could lead to platforms "placing themselves in a gatekeeper position which may easily translate into (dominant) positions with strong market power". There is currently significant public concern about the potential abuses of market power from data-driven digital platforms.
2. **Conglomerate effects** – the desire of, and rationale for, companies with data-driven business models that are successful in one service to expand into unrelated services is a normal commercial freedom. There may be instances where the data collected and insights derived can be used to gain advantage when competing in the new area that goes beyond normal integration of adjacent areas, such that other competing companies are not able to replicate. Equally there are examples of companies making acquisitions that appear separate from their core service, and so do not raise competition concerns since market shares on a traditional market definition are unaffected, but where the ability to collect and aggregate data (about users, customers or more generalised insights about consumer behaviours) confers a powerful competitive advantage. Competition authorities are becoming more alert to such situations but have not typically defined new tests to evaluate them.
3. **Privacy and trust** – many data-driven business models raise potential concerns about privacy and trust. There are differences by geography and demographic segment but here too public sensitivity is rising quickly as seen in recent events. This report does not address this broad and rapidly evolving topic directly[4], but privacy regulations, and user attitudes to trust, will inevitably shape the competitive landscape in a significant way.

The report discusses concerns that traditional competition policy focuses only on price effects when evaluating consumer welfare. As a result, current approaches might not address other, important aspects, including the potential longer-term negative impacts of a heavily integrated and concentrated data value chain. There are areas of the emerging data value chain that could function more effectively where the unique characteristics of data and data-driven businesses could benefit from an adaptation of policy approaches. There is a need for more careful inspection of the implications of the power of data, open innovation and the prospect of market entry, in addition to a traditional consumer benefit perspective. As an example, an important priority for policymakers would be to foster competition across digital ecosystems by eliminating regulatory barriers that restrict companies from competing openly in the data value chain.

2. 2.20 billion monthly active users as of March 31, 2018, source: newsroom.fb.com
3. European Commission, Business-to-Business relations in the online platform environment, Final Report 22nd May 2017
4. Other GSMA publications discuss this topic, including Safety, privacy and security across the mobile ecosystem (2017)

- From a business perspective, there is a clear need for consistent rules to govern all players who operate based on similar data, regardless of the technology or infrastructure they may use or services they may deliver. As traditional industry segmentations disappear, there is an opportunity for greater competition but regulations targeted at specific players can create bottlenecks in the market and distort competition. Sector-specific regulation in new markets should generally be revised to be consistent for all competitors with access to user data. For example, legacy regulatory barriers to mobile operators' full participation in data markets should be removed to foster maximum competition in the data economy.
- Direct and indirect network effects are an inherent factor in many data-driven business models, so policymakers need access to up to date research and tools so that they can evaluate these effects and their potential role in creating market power. Such power may lead to dominance in one market but also a formidable competitive position across distinct markets.
- Transparency regarding access to data is an important ingredient for all data-driven businesses. It is right that companies that invest in the means to collect data in accordance with prevailing law should be allowed to then use it to earn a commercial return - otherwise the incentive to invest is removed - but this needs to be done in a way which is understood by users and not in a manner which could be considered surreptitious or underhand, which could thus undermine the trust needed to allow data driven models to work and develop.

The value of data as a strategic asset and powerful source of economic value is clear. The commercial success (and in some cases mass popular appeal) of data-driven businesses attests to the impact they are having around the world. The rapid adoption of all forms of online services demonstrates the consumer demand for such services with resulting benefits in terms of consumer choice and lower prices for all forms of services: from entertainment, to travel, to communications. These benefits should not be taken for granted or assumed to have been attained without broader impacts on consumers and competition, as is increasingly recognised in public debate. Close attention is required to ensure the markets for data and data-driven services continue to operate as efficiently as possible, that consumers are afforded an adequate level of protection, and that competition policy evolves to take account of the specific characteristics of data driven businesses and their impact on the competitive dynamics of the data value chain.

Introduction

The purpose of this report is to understand the data value chain in terms of the activities and business models involved in collecting and monetising data. The report additionally considers the role that mobile network operators (MNOs) are taking within the value chain and the issues and barriers they face when seeking to develop and offer services in these markets. Our objective was to take a holistic view, both of the sources of data and of the ways in which data is processed/analysed and used for commercial purposes[5]. While written on behalf of the global mobile operator industry association, GSMA, the perspective is wider and covers all industry sectors.

Most organisations have for a long time recorded details of their customers and transactions. What has changed recently is that these interactions are increasingly digital, with data being generated at a much higher rate due to the development of the internet, smartphones and mobile broadband. Rapid increases in the scale of computer processing power and the ability to combine highly disparate data sets at a much greater level of detail massively expand the potential for data-driven business. These attributes are commonly referred to as the ‘four Vs’: volume, variety, veracity and velocity, all of which are increasing. Further, personal information and preferences have become valuable products in their own right, thanks in large part to the information garnered on social media platforms and the ability to use them for the micro-targeting of advertising. Internet of Things technologies are having a similarly significant impact on the generation and capture of operational data within organisations and their supply chains.

IDC[6] predicts that the “digital universe” (the data created and copied every year) will reach 180 zettabytes in 2025, up from less than 10 zettabytes in 2015 (a zettabyte is the equivalent to 1 billion terabytes, or approximately the capacity of 1 billion typical laptop

5. As a stylistic matter, the report refers to “data” as a business phenomenon in the singular form
6. Taken from Haber, P., Lampoltshammer, T., Mayr, M., (2017) Data Science – Analytics and Applications

hard-drives). New business models are evolving to capitalize on this. IDC estimates that big data and business analytics revenues in 2016 were $130 billion and are expected to exceed $200 billion by 2020. The rapid volume growth is increasingly driven by a wider variety of connected devices from thermostats to connected cars or body sensors. The mobile industry has been key to enabling what some have termed 'the fourth industrial revolution'[7], in which technology becomes ever more pervasive and integral to aspects of daily life, as the cost of collecting, transmitting and storing massive amounts of data falls rapidly. Of course, some traditional economic sectors have incurred significant disruption in the face of competitors with data-driven models.

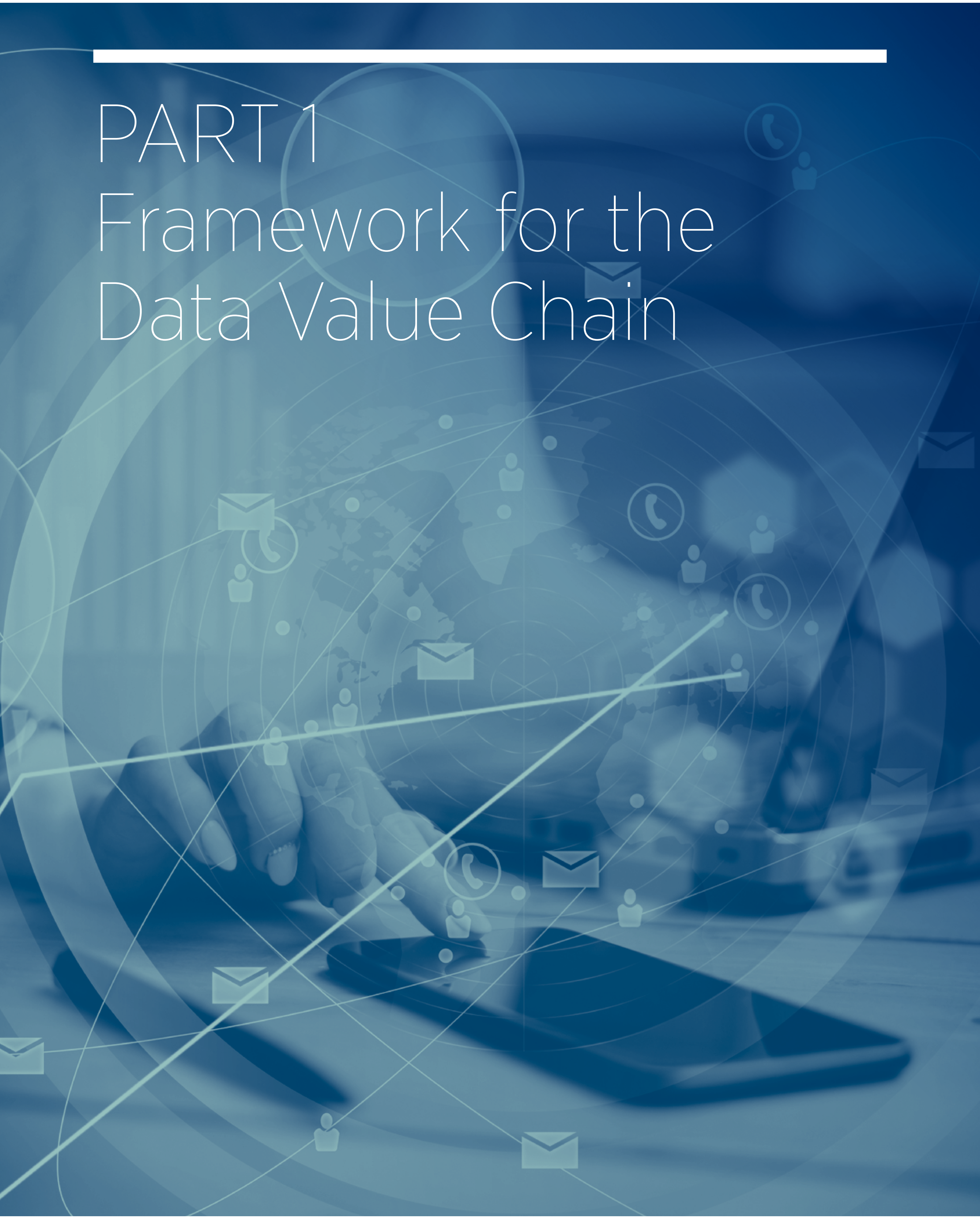
The report discusses the different types of data (including personal and non-personal), business models and industry players. It considers how value is being created at various stages of the value chain and what is the role of mobile operators in the data value chain. Previous reports on the internet value chain or mobile economy have succeeded in making quantitative estimates of the impact of these vital technologies, but the data value chain is not yet mature nor transparent enough and so this report relies primarily on qualitative observation and analysis, drawing also on other published research. It also considers how well the market is functioning in a general sense. It is possible to identify areas of potential or actual market distortion, which is not unusual for a major technology and business model disruption where the market and technology realities have evolved rapidly and regulatory approaches have struggled to adapt at the same pace.

In the final section of the report we therefore discuss potential implications for public policy. In some cases it is too early to conclude where policy changes or interventions are needed. There have already been cases of competition policy intervention and plenty of debate on the right balance between encouraging innovation and enforcing openness and competitiveness in data markets. It seems clear that there are instances where policymakers should foster fair and efficient competition across digital ecosystems by eliminating regulatory barriers that restrict companies from competing in the data economy. There are of course wider policy implications in the rise of the data value chain, such as the need to redesign policies on taxation, education and skills, as well as the need to modernise policies and practices on privacy and security – the latter having been addressed by a GSMA report earlier in 2017[8].

While produced on behalf of the GSMA, following close dialogue with a number of its members, this report is independent and does not necessarily represent the views of the association or its members. It was produced as a contribution to an important public debate and neither A.T. Kearney nor the GSMA accept responsibility for any other use.

7. "The Fourth Industrial Revolution", Klaus Schwab, founder and executive chairman of the World Economic Forum
8. GSMA Report, Safety, privacy and security across the mobile ecosystem, February 2017

PART 1
Framework for the Data Value Chain



Types and Characteristics of Data

Data is a unique economic good. Many associate data with abundance but this is misleading. Instead, the matter is one of variety – in fact that there are enormous numbers of scarce or even unique pieces of data. As a form of intangible asset, it shares characteristics with several other kinds of capital good, but combines them into a distinctive mix unlike any other asset.

Firstly, data has a lack of fungibility, that is to say, no two pieces are the same or can be substituted for each other. There are many streams of information, each is different and it is hard to compare and value them in a relative way. This makes it difficult for buyers to find a specific dataset and to put a price on it in a way that is comparable with other data.

Secondly, it is non-rivalrous. This means there is the ability for data to be used by more than one person (or algorithm) at a time and it is not consumed in the process. Any given piece of data can easily be used for a multiplicity of purposes independently. For example, readings from a weather station could be provided to weather forecasting agencies to inform current and imminent weather conditions and also to scientists studying longer term weather trends such as climate change. Neither use impacts or detracts from the value to the other users. This makes data similar to other intangible assets such as intellectual property, where for example, patents can be licensed to multiple users for a variety of uses (provided a licensee did not request and pay for exclusivity) and the price is based on the input and no value ascribed to the resulting outcome.

Finally, data is known as an experience good. This means its value is only realized after it has been put to use and it has no intrinsic value of its own. Other examples of experience goods are films and books, where there is a degree of uncertainty around whether it is worth investing your time, money and attention in them but the only way to find out is to try them. While all data has the general characteristics described above, there are many ways in which data can be classified. Figure 1[9] below highlights some of the key dimensions that are relevant for our analysis.

Figure 1

Key dimensions and data types

Dimensions	Data Types		
Personal	Volunteered	Observed	Inferred
Personal	Private		Public
Personal	Identified		Pseudonymised
Non-Personal	Anonymous		Machine Data
Timeliness	Instant/Live		Historic
Format	Structured		Unstructured

9. There is a subset of machine data which could be considered Personal. In this context we refer to machine data that cannot be used to reasonably identify an individual

While many of these characteristics can be applied to any dataset, there are a few which have a direct bearing on the value of the data and how it can be used.

Personal Data

Personal data is anything pertaining to an identified or identifiable individual and may be private or public. Companies have always collected private personal data such as name, address, often bank details. Health records would be another example of private personal data. Companies and bodies that collect this data have clear responsibilities to store it safely and to ensure it is used only for the intended purposes. In most jurisdictions there are also clear guidelines on what personal data can be collected and how long it can be stored for.

However, there are grey areas. A smart thermostat that learns the behaviours of a household in order to find the best time to activate the heating is machine generated data in the first instance but clearly the behaviours being recorded are personal, in this case sleeping habits.

A recent World Economic Forum[10] paper focusing on personal data proposed that data could be grouped by how it was created or shared:

- **Volunteered** – data and content explicitly shared by the data originator, e.g. social medial post, personal information given to service providers such as banks, uploaded photographs, measurements from connected devices etc.
- **Observed** – data which is collected in a non-intrusive way about the context or behaviour of a subject (with or without the subject's knowledge that they are being observed) which could include location or device used to share content, time taken to carry out a task (e.g. to make an online purchase), other services used etc
- **Inferred** – these are new insights which are created by analysing and processing volunteered and observed data, and as such could not be considered primary facts and depend on probabilities, correlations, predictions etc. Examples include credit rating, insurance risks, socio-economic banding, search result or 'next purchase' suggestions

A more recent development in this area is that of pseudonymous data. This is defined as personal data that has been subjected to technological measures such as hashing or encryption such that it no longer directly identifies an individual without the use of additional information. This is particularly relevant in Europe as it is referred to within the General Data Protection Regulation[11] (GDPR). Pseudonymous data is still considered a type of personal data. Under the GDPR pseudonymisation is a safeguard that reduces privacy risks. Data controllers applying it are allowed to process data in new ways, as long as the processing is compatible with the purposes for which the data was originally collected. However, the onus is still on the data controller to ensure that the data is genuinely pseudonymous and cannot be used to re-identify the originator.

Besides private data, there is also personal data that is available more publicly, for example social media posts, public registers and, in some countries like Norway, citizens' income and tax payments. In the case of observed data collected about individuals, it is not always clear where the boundaries of how this can be used lie. For example, as facial recognition technology becomes more accurate and ultra-high-definition cameras more widely available there are increasing concerns that programs could easily make matches of individuals from CCTV images given the proliferation of private CCTV images and public social media postings. This matters because increasingly individuals could be identified and this information can be amalgamated with others to build up a detailed picture of a person's movements or behaviour. Police and security agencies are increasingly making use of this as part of crowd control and anti-terrorism actions but the applications in the commercial space are large too.

10. http://reports.weforum.org/rethinking-personal-data/near-term-priorities-for-strengthening-trust/ - Accessed 26 October 2017
11. GDPR is a regulation being introduced in the European Union (effective from May 2018) intended to strengthen data protection rights for individuals. The regulation explicitly recommends pseudonymization of personal data as one of several ways to reduce risks and make it easier for controllers to process personal data beyond the original personal data collection purposes or to process personal data for scientific and other purposes.

Non-Personal Data

There is a lot of data which is machine or system generated which is clearly non-personal, e.g. engine performance data, transport ticket machine sales, energy usage data, stock price and transaction data, blockchain records etc. Data relating to individuals but which is fully anonymised is also considered to be non-personal, e.g. blockchain ledgers. In fact, almost every business and government agency are probably generating large volumes of non-personal data, for example daily sales ledgers, transaction numbers, shipping records etc.

Timeliness

The time dimension of data is an important characteristic. Some data is a specific measurement recorded at a point in time. For many applications, such as news feeds, price points, stock market trades etc., the timeliness of data is crucial as they require real-time stream processing. The more up to date the better and the value of the information declines with time, sometimes measured in milliseconds. In financial trading, there are examples of firms building direct fibre links between facilities or trading platforms to minimise any delay in data transmission. There are other instances where a user chooses to pay a premium in order to be able to get access to data earlier than others. These include the latest market research on shopping trends, because it is important to their business and competitive behaviour. Other forms of data may have a more dynamic nature, such as traffic information applications, where it is the change over time that is important (e.g. speed and traffic flow data) and a company needs to have the infrastructure to collect and process the data streams on a continuous basis to be able to generate useful services based on these.

Structure

Another characteristic which affects the value of data is whether the data is structured or not. Structured data comprises clearly defined data types, organised into searchable fields whose pattern makes them easy to interrogate, for example a set of readings from a particular sensor, or a database of customer records. Unstructured data, on the other hand, comprises data in formats like audio, video and social medial postings, which are much harder to organise and search. An often-quoted rule of thumb within the data industry is that unstructured data accounts for more than 80% of all data.

However, as described in the data processing step above, there are many companies developing services that enable unstructured data to be organised, indexed and analysed in ways that then create value from it. There are several challenges around unstructured data:

- **Relevance** – often there is a lack of insight from the data. Machine learning is often applied to unstructured data and can be very powerful and surprisingly accurate but it is certainly not infallible. There is a need for some awareness of the trends in order to train a computer (correlation or causation issues). As more value is derived from inferred data and predictive models, for example with health insurance quotations, the consequences of the cases where the inference is wrong, however small, need to be considered.
- **Volume** – the speed at which the volume of unstructured data is growing is extremely rapid for some businesses to keep up with. This presents challenges for using and securing the data. It also requires an infrastructure which many businesses don't have or can't afford.
- **Quality** – a large volume of unstructured data is collected at source but remains unverified (e.g. the sensor could be inaccurate, the time stamp may be wrong etc.) and this fact could be lost in the subsequent processing and use of the data and there need to sufficient checks in place to verify the source data and its accuracy. In cases such as using inputs to update credit references there is a structured process to minimise such cases, but in other cases there are less checks. For example, the consequence of a mistyped search could be a stream of adverts for irrelevant or inappropriate products.
- **Usability** – for unstructured data to be usable, businesses will have to come up with a way to locate, extract, organize, and store the data. This means coming up with an entirely new type of database to store information that does not fit the mould and there are companies developing software and services to solve this issue for enterprises.

Data Value Chain Framework

In terms of understanding any value chain, it is instructive to look at the individual activities/steps, assess how value is created at each step and what types of business models are employed to do so. In the data value chain, many companies operate across multiple individual activities to differing extents, but a review of the discrete activities is helpful to provide insight into which activities generate value and also where there may be barriers to value creation.

This section describes a framework to assess the activities that develop data into assets which companies can exploit to generate value as part of their commercial activities. We have called this the data value chain. Previous studies have developed similar frameworks. The OECD[12] in 2013 and the European Commission in 2014, with the latter presenting its own data value chain[13] to argue that data needs to be made interoperable as part of a data economy. Academics such as Miller and Mork (2013)[14] and Curry (2016)[15] have defined their own taxonomies and we have drawn on these where applicable, as well as Tang's model (2016)[16] specific to data analytics. The framework described in this report focuses on the commercial activities by which value is created and consists of four high-level stages with sub-activities.

Figure 2

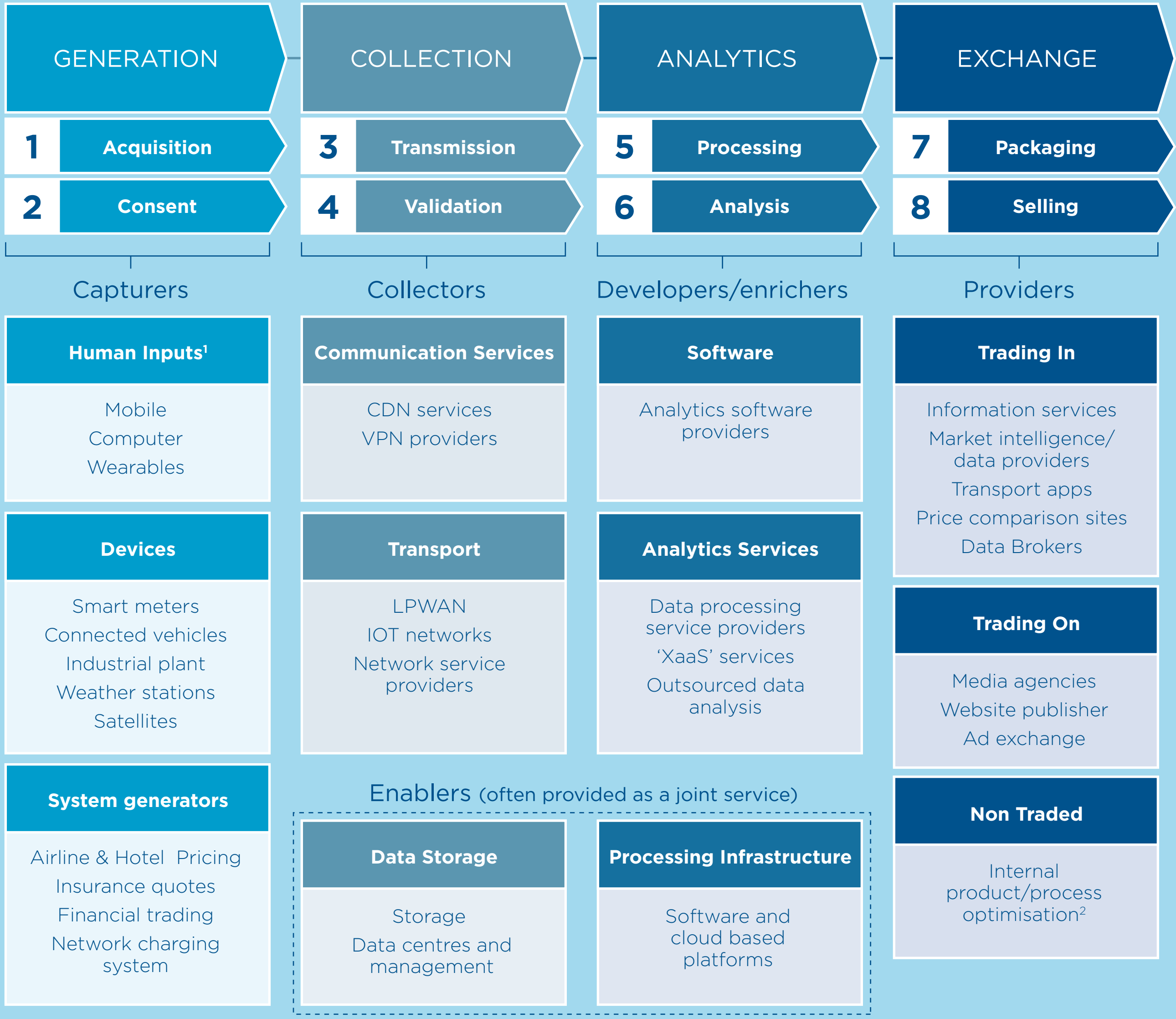
Data Value Chain framework

The value chain begins of course with data generation, which is the capture of information in a digital format. The second stage is the transmission and consolidation of multiple sources of data. This also allows for the testing and checking of the data accuracy before integration into an intelligible dataset for which we have adopted the term collection. The next stage is data analytics, which involves the discovery, interpretation and communication of meaningful patterns in the data. The final stage takes the output of these analytics and trades this with an end-user (which may be an internal customer of a large organisation processing its own data). Unlike most value chains, the data is not consumed by the end-user, but may be used and then reused or repurposed, perhaps several times, at least until the data becomes outdated. Even then it may become part of a historical trend which still has value. We therefore use the term exchange for this.

12. OECD (2013), "Exploring the Economics of Personal Data: A Survey of Methodologies for Measuring Monetary Value",
13. European Commission report (2014) A European strategy on the data value chain
14. Miller, H & Mork, Peter. (2013). From Data to Decisions: A Value Chain for Big Data. IT Professional. 15. 57-59. 10.1109/MITP.2013.11
15. Curry E. (2016) The Big Data Value Chain: Definitions, Concepts, and Theoretical Approaches. In: Cavanillas J., Curry E., Wahlster W. (eds) New Horizons for a Data-Driven Economy. Springer, Cham
16. Tang, C. (2016) The Business and Economics of Information and Big Data

Figure 3

Data Value Chain structure and activities

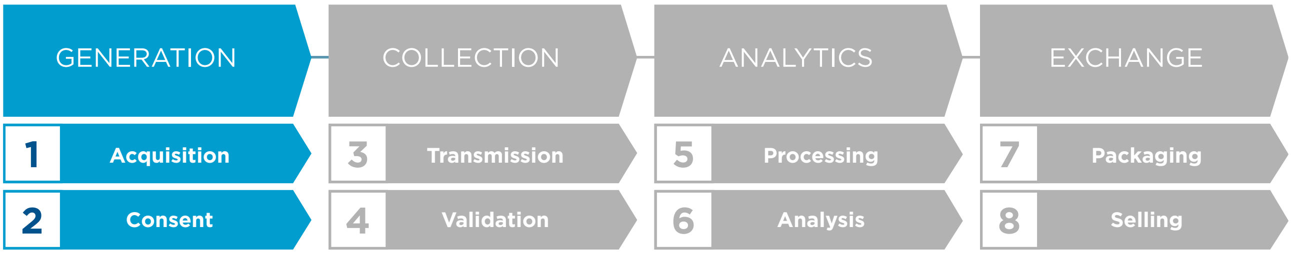


1. i.e. a person directly generates data and inputs via mobile, wearable etc
2. i.e. the organisation collects and analyses data for its own (value-creating) activity only

Source: A.T. Kearney

These four principal stages are further illustrated in figure 3 above, with examples of the types of services carried out or used at each stage. The individual activities described below.

1. Generation



1. Acquisition

Clearly the first step in the value chain is acquiring the data. The actual source of the data may vary: data may be internally generated or obtained externally; may be passively or actively created by a human, a sensor or a system, may be structured or unstructured in form.

In one of the most intuitive cases, data is created every time a user posts an update on a social media platform. Facebook has over two billion users, sending 31.3 million messages every second, with 2.7 billion likes tracked daily. Typically, it is possible to capture a range of other data beyond the content of the post, including potentially where the user was when he/she posted or liked, what other use was being made of the device before and after, and so forth. A few keystrokes represent a significant volume of data and the repetition over time represents a significant dataset from which value can be extracted.

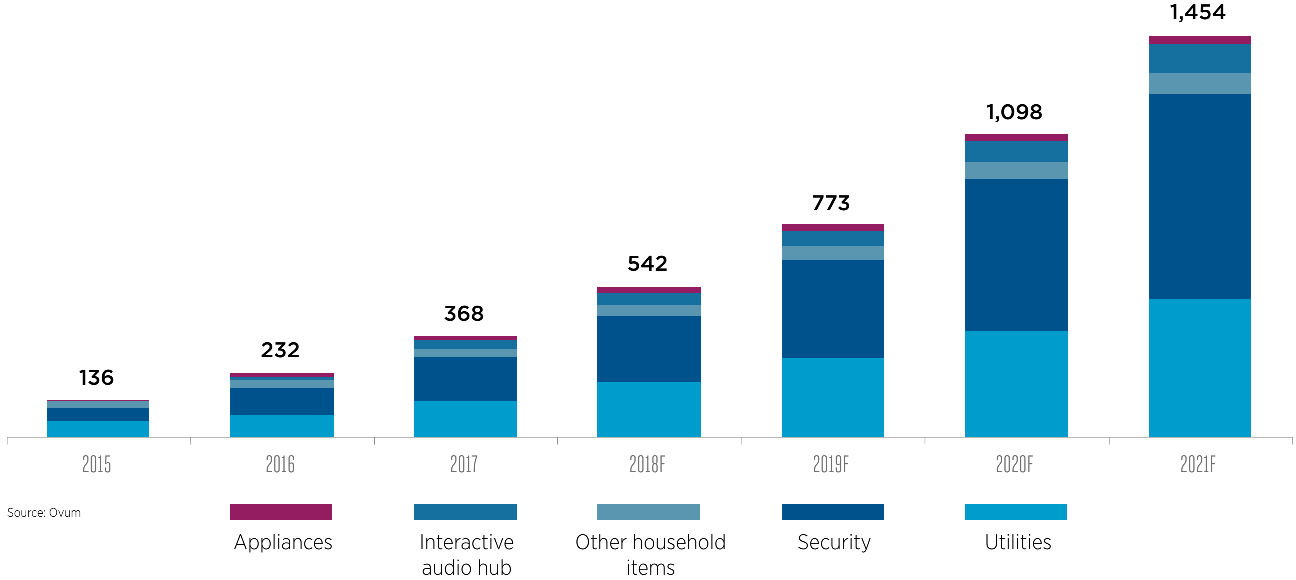
Transactions between individuals but also organisations and machines produce equally significant amounts of data. Large retailers and B2B companies generate data on a regular basis about transactions which consist of product IDs, prices, payment information, manufacturer and distributor data, and much more. Every time a passenger enters a mass transit system, or checks an arrival time, this creates a data record of activity and interest which at an aggregate level is a powerful source of insight.

The sophistication of sensor and recording devices to provide real-time data is another driver of the rapid growth in data generation. Weather stations automatically measure, record, and report a standard set of date including temperature, wind speed and rainfall on a regular and periodic basis. At CERN, the largest particle physics research centre in the world, the Large Hadron Collider (LHC) generates 40 terabytes of data every second during experiments. More commercial uses of machine data are those used to monitor machinery and constantly evaluate operational performance, be this a jet engine, an energy system, or the performance of the server farm where the data is stored.

Smart Homes devices are another source of data and the market for such devices is forecast to experience rapid growth as the 'Internet of Things' drives a wide range of connected devices that are able to sense and measure activities and then connect to services via the internet. Figure 4 shows the expected growth in Smart Home devices and the growth especially in security and utilities devices, such as security cameras visible via a smartphone app and connected utility meters which report usage readings and have the potential to respond to dynamic energy tariffs.

Figure 4

Unit Sales of Smart Home Devices (millions)

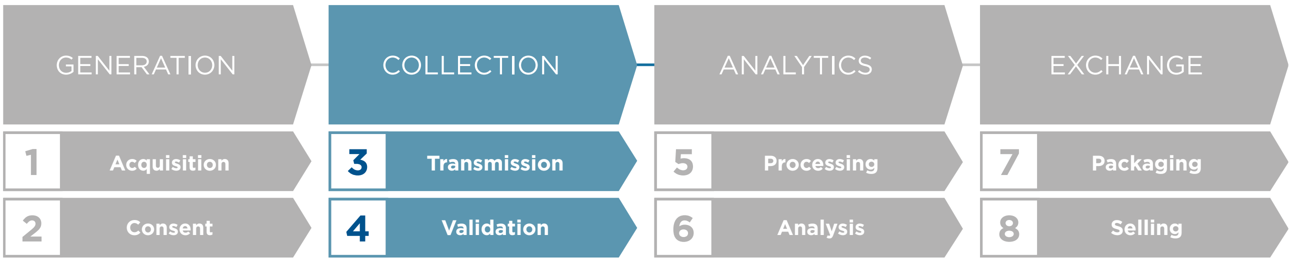


Source: Ovum

A further source of data is that which is system-generated using algorithms that employ a set of input parameters to create specific new data points which can be considered as primary data. As an example, seat prices for most commercial flights are generated in conjunction with a specialised software which takes many inputs to create a specific price for each available space on each flight, e.g. route, time and date, class of service, ticket conditions, advance purchase period, whether connecting or direct, etc. and then dynamically updates them to respond to demand as seats are sold. Financial trading algorithms are used extensively to make trades for equities and other instruments and in the process, determine share prices that vary constantly. The output of these systems may ex post represent a transaction history of flights or hotel rooms booked and shares traded, but in real time they have in fact generated discrete data points about a specific item or event, which when combined afterwards then underpins those transaction histories.

2. Consent
In most cases it is at the generation stage that the service provider needs to gain an acceptable basis for processing the data. There are several ways they can do this, including gaining explicit user consent, performing duties as part of an agreed contract or legitimate interest conditions. For machines and system-generated sources this should be relatively unproblematic in terms of consent and access[17] as it will either be company-internal or will be covered by the broader commercial agreement related to the supply of the monitoring equipment and set-up, e.g. smart utility meters, factory production monitoring sensors etc. However, it is important to know that there will be some grey areas such as data generated by operating systems of user devices or the various tracking apps used by websites to monitor user interactions. For data generation based on human inputs in many cases, the collection of data at source is based on an implicit transaction, where a consumer uses a free online service and (as part of the terms of service when first registering) agrees to their personal data being collected and used by the provider. The most obvious is that of digital advertising-based business models, which allow consumers to use popular services for free, for example search engines or social media platforms, while using the information collected to sell advertising on a targeted basis.

2. Collection



3. Transmission
At the collection stage the data must be transmitted from its capture point to where it is being stored using some type of networking infrastructure. The available options include telecommunications networks and public internet services (including Wi-Fi access). There are also dedicated (or special purpose) networks such as private radio infrastructure, Low Power WAN networks being developed for Internet of Things applications, (e.g. SigFox), and SCADA (Supervisory Control And Data Acquisition) networks used in manufacturing plants. In some instances, the transport service is enhanced by a specialised service such as Content Distribution Network (CDN) or Virtual Private Network to enhance reliability of delivery and end to end security.

4. Validation
As part of the data collection process, primary data is also pooled with other associated data (from other locations, sources or time periods for example) and will often undergo validation to ensure that the data received is correct, that it comes from a verified source or user, that the sensor providing it is functioning correctly. Readings from smart utility meters are validated in this way. Once the data has been transmitted it is likely to be combined with other data of a similar type. This increases the scale, scope and frequency of data available for analysis. For example, temperature sensor data may be combined with historical data from the same sensor or pooled with data collected at the same time from other geographical locations. This then forms a raw dataset where players specializing in collection will validate the data.

17. In Europe in the future, it might maybe more complex as the proposed ePrivacy regulation, still under discussion, includes M2M in its scope.

The industrial Internet of Things provides an effective example. Businesses operating in this area often provide asset performance management and operations optimisation by enabling industrial-scale analytics to connect machines, data and people. An example from the aviation industry is the monitoring of aircraft landing equipment. On-board sensors that detect hydraulic pressure and brake temperature generate the data. The data is then communicated to a Quick Access Recorder – an onboard industrial device for data ingestion and storage. If connectivity is available from the aircraft to ground, the data is sent to the airline's operations hub in real time; if it is not available the data is copied from the Onboard Monitoring Computer to the cloud after landing using a USB flash drive. The data is stored and anomalies are diagnosed, using historical data for reference, and fixes applied where needed. In subsequent stages the actual sensor data is continuously compared with predicted data based on the digital twin. For example, actual brake pad temperature is compared against the brake pad temperature threshold. This determines the remaining useful life for each subsystem of each piece of landing gear and so informs the maintenance schedule for that plane.

Storage and Processing Infrastructure

There are two further key enabling services within the value chain which support the activities handling the actual data once it has been collected. In our framework they are formal enablers of the collection and analytics steps of the value chain.

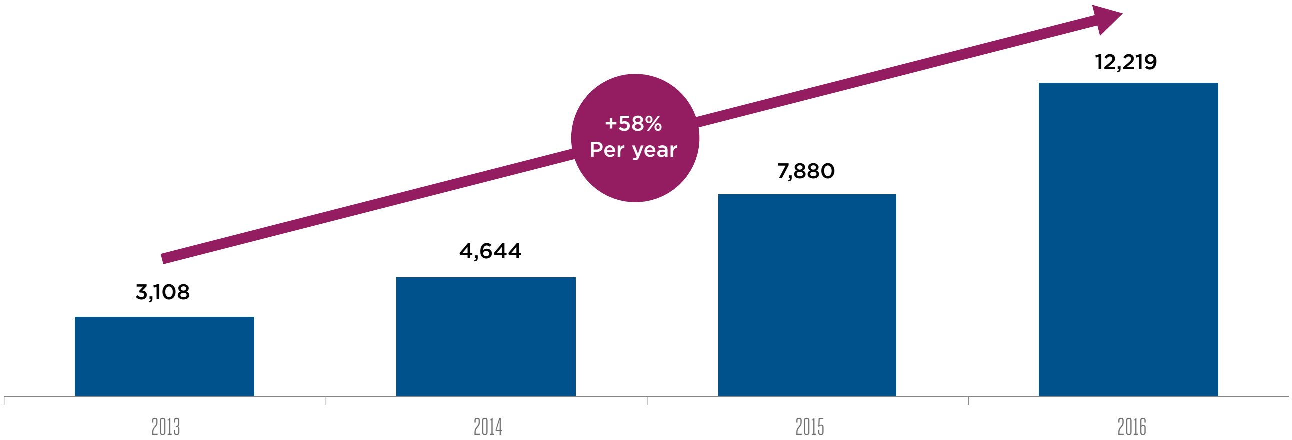
Data storage is the activity of organising the collected information in a convenient way for fast access and analysis. Generally, this is carried out using hardware in data centres with multiple software tools available to catalogue and index data so that it can be easily accessed, backed-up, encrypted and updated. In its simplest form, this could take place on an office server. In practice, the service is massively scaled up to fill aircraft-hanger sized data centres with vast storage capabilities. The advantage is that the unit cost of storing and securing data is greatly reduced. Upgrades to the underlying hardware can also easily, and seamlessly, be carried out as new technologies are developed that enable greater and cheaper storage capacity. Such services can be provided in a dedicated physical site, often with a mirrored site for resilience and disaster recovery purposes, but increasingly are provided in networks of sites, commonly referred to as cloud storage, where a given dataset may be distributed across the storage network, rather than residing in a dedicated physical space. This further increases the resilience and improves the economics.

Building on these storage services, processing infrastructure is very often bundled as offered as part of the same service and the raw data processing power can be provided and scaled in a similar way. The basic building block is a processor such as those found in a laptop but hugely scaled up and provided in an environment where the processing infrastructure, down to the actual processor microchips, are effectively shared, with multiple applications running at the same time and able to load share demands. Whereas the processor on a typical laptop will be effectively idle for large periods of time (even when turned-on), in a shared environment it can be repurposed to carry out other tasks for other users in this idle time. This delivers very large gains in the processing power available and the economics, making large computing power available to small users who previously would not have had the resources to access such capabilities.

The success of this model, at least in terms of demand, can be seen in the revenues of one of the leading providers of cloud-based storage and processing infrastructure, Amazon Web Services, as shown below in figure 5.

Figure 5

Revenues of Amazon Web Services, $m



Source: Capital IQ

Such a rapid pace of growth is driven by the large-scale shift of storage and processing activities from proprietary, own-use (and often in-house) data centres to cloud based services and being seen across all industry segments, including public sector, with an increasing proportion of government services delivered from cloud-based infrastructure.

3. Analytics

The analytics stage is where the data is processed and put to use to generate insights and useful information. Different frameworks have been developed which break this stage in multiple sub-activities (for example see Tang 2016) which explain the technical steps involved. From a commercial viewpoint it makes sense to think in terms of the operational processing of data (which could be a series of episodic activities or a continuous process), and the analysis which is the structuring and 'intelligent thought' of making sense of the outputs and drawing insightful conclusions.

5. Processing

In order for analysis to take place, data needs to be in a suitable format. Processing converts the data to a format appropriate for mining and analysis and guarantees the integration of only the effective and useful results and so converts raw data through a series of steps into information and knowledge. For example, a series of temperature readings over a time series would be raw data. Putting them together and identifying a pattern over the period would be information. Without this step there can often be problems as network bandwidth and/or data centre resources are swamped by capacity demands and this can cause potential processing delays. Placing the systems capable of performing analytics at the network's edge can lessen the burden on core network and IT infrastructure.

As part of the processing step, data may undergo mining. This involves using sophisticated mathematical algorithms to segment the data and establish links between pieces of data. The key features of this include automatic discovery of patterns, prediction of likely outcomes, and the creation of actionable information with a focus on large data sets. Examples of this include interrogating customer purchase information and identifying patterns that show the likelihood that a customer that buys product A will also buy products B, C & D and also include inputs such as age, socio-economic grouping, store locations to establish clearer pictures of typical purchase groupings and so inform targeted marketing activities and offers.

As an example, the fashion industry is one industry which is embracing data mining. Data analytics is not new to this industry, it has for a long time analysed sales information to direct production and identify future designs. However, the growth of unstructured data, especially that acquired on mobile devices or social media sites, has opened up many more analytic opportunities. This data may be made up of images, audio, videos and text. Apparel retailers are using cognitive computing to understand more about what customers might want to buy and use this insight to influence designs for future seasons. This technique relies on data mining to recognize patterns and uses this insight to simulate the human thought process. As well as giving more insight into trend prediction, consumer preference data can be analysed much more quickly. This can enable two weeks additional selling time over competitors which can be decisive in the highly competitive environment of fashion.

6. Analysis

At this point it is necessary to develop hypotheses around the content of the data by taking the outputs of the processing and employing them to perform such tasks as making predictions, understanding actions and behaviours. A key determinant as to whether the output of the analytics constitutes a meaningful insight rather than a random correlation is based on the need for both functional and technical expertise at the hypothesis and analytics stage.

The betting industry[18] is one of many which is being transformed by data analytics. New offerings provide a good example of how hypotheses are developed from data mining insights and then tested. In the future, punters will be able to place bets with a bookmaker as races are being run. Also available to gamblers will be detailed information on sectional times within races, precisely how much ground each horse has covered, its stride length and pattern, and how quickly it clears an obstacle over jumps. The positional fix which allows for these measures comes from a satellite, via a tag located in the horse's saddlecloth which reports to the data hub in real time. Gamblers could predict through their bets a horse's overall performance, rather than simply if it will win the race or not, with the breadth of data points potentially reducing the scope for fraudulent manipulation of the outcome. Overtime the patterns that emerge from the success of bets linked to the data collected on a horse's performance will offer further insight to refine the algorithms generating the odds.

There are many ways in which the analysis can be designed and conducted. A simple approach would be to define the mathematical operations and statistical handling, as in the temperature example defining the period to analyse over and how to combine and weight readings, and to then design the outputs, most likely in the form of charts or graphs. In practice, much more complex approaches are applied to pull together data from previously unconnected sources to develop new insights. Machine learning uses advanced techniques so that rather than repeating the same algorithm, the algorithm can actually improve as it performs more calculations and so refine the outputs based on feedback. This is especially effective in situations where there is a range of outcomes and a level of ambiguity in the input data, for example voice and facial recognition systems. These techniques have enabled the development of applications which anticipate human actions such as software which prevents emails being sent to the wrong recipient (see callout box).

Machine learning uses data mining techniques and other learning algorithms to build models of what is happening within a data environment so that it can predict future outcomes. These techniques have enabled the development of a software which prevents emails being sent to the wrong recipient. The software is customised for a particular client – this is often legal or other professional services firms where the content of outgoing emails can be highly sensitive. Initially, there is the collation of millions of data points from across the entire email network of an organisation. From this mass of unstructured information, the technology categorises and maps key data relationships. Subsequently data science and machine learning algorithms are applied to detect patterns of behaviour and anomalies across the network to develop an understanding of typical sending patterns and behaviours. With this knowledge, the software can analyse outgoing emails and detect any anomalies. If an anomaly is detected the software will temporarily stop the email from being sent and display a notification to the sender. The company behind the technology monetise their algorithms by selling the software licences.

18. The Guardian - Bookmakers to embrace 'big data' to shift racing's betting landscape – Published 6 October 2017

Artificial Intelligence is a fast-growing type of analysis in which, rather than repeating a predefined algorithm, the software adapts its approach based on the findings and so over time can increase the accuracy and depth of insight of outputs. Deep learning is one such way that these capabilities may be developed. This a subset of machine learning and is concerned with developing algorithms that teach computers to learn by example rather than structured instruction programming. The computers use various mathematical and statistical models to achieve this.

In all aspects of analytics it is important that the data input matches the purpose and the analysis is carried out with knowledge of the context. A US research project to try to predict unemployment figures by tracking the correlation between employment figures and social media search terms such as ‘unemployment’, ‘situation vacant’ and ‘jobs’ spotted a large spike in their data that could have been a precursor to a rise in unemployment. With further (human) reflection, it turned out the spike was due to trail of social media posts following the death of Steve Jobs. This supports Schroeder (2016)[19] who states that valid inferences can only be reliably drawn from data when the analyst has a thorough understanding of the data and the context from which it was drawn, and that this can be lost as datasets are increasingly combined and aggregated and not taken from primary sources where the context is clear.

Examples

Facial Recognition – This is being used for example to enable facial recognition to be used as a security tool. Unlike a conventional log-in method, this technology confirms that the face is actually that of the user. Facial recognition systems use statistical algorithms to translate an image into a string of numbers that describe a face. When comparing two different images of the same person, the algorithm will return similar outputs (i.e. closer numbers) for both images, whereas when images of two different people, would return very different outputs.

Predictive Policing – Data insights and analytics are also being applied to policing. Trend analysis has been used, most notably by Chicago, to analyse trends and patterns in crimes and locations and use this to create heatmaps of where future crimes may occur. It can do this by processing inputs including arrest reports, property reports, commercial databases, web searches and social media postings to create ‘threat scores’ which are more predictive than just using past behaviour. As observed by the US Government, when designed and deployed carefully, data-based methodologies can help law enforcement make decisions based on factors and variables that empirically correlate with risk, rather than on flawed human instincts and prejudices. However, it is important that data and algorithmic systems not be used in ways that exacerbate unwarranted disparities in the criminal justice system. For example, unadjusted data could entrench rather than ameliorate documented racial disparities where they already exist, such as in traffic stops and drug arrest rates[20].

19. Big data business models: Challenges and opportunities, Ralph Schroeder, Cogent Social Sciences (2016), 2: 1166924.
20. https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf

4. Exchange

GENERATION	COLLECTION	ANALYTICS	EXCHANGE
1 Acquisition	3 Transmission	5 Processing	7 Packaging
2 Consent	4 Validation	6 Analysis	8 Selling

The final step in the value chain is putting the insights and knowledge generated to commercial use in order to realise the value created. Packaging and selling are the two sub-stages of the exchange step. Not all these steps will occur in each case or in the same way, depending on how these insights are put to use. For example, a company using insights about its customer needs to refine its own commercial offer will not package the insights in the commercial sense, but the insights team will distil key findings into a presentable format that product development teams can act on.

7. Packaging

This step expands the usefulness of outputs from the analytics stage, by taking insights and putting a 'wrap' around them to create a product which is accessible and usable for a potential end user. In many cases this will have a technology component as the information is delivered via an automated channel, for example APIs (Application Programming Interface) that allow software programmes to share information in a standardised format.

There is an increasing desire to communicate insights drawn from data in a compelling fashion. The use of visual tools, 3D models and effects is an area of innovation that is creating new ways for users to see results and even interact with them. These visualizations can become an output in themselves or may be a sub-stage of analysis since they creatively present the data allowing users to develop new hypotheses to test and inform future analysis.

An example of this is provided by credit rating agencies. These organisations gather vast quantities of information about individuals from lenders, aggregate it and use it for data analytics. They then package up the insights obtained on consumer banking and spending habits and sell bulk lists to banks and credit card companies and also create the tools and interfaces that allow financial service companies to carry out credit checks on specific individuals, for example when applying for a credit card or mortgage and then charge for each enquiry made.

8. Selling (or own use)

The final step is the actual trade or use of the data and insights. This is the main step where the cumulative value derived from the previous steps is realised. Value is added at this stage using data technology as well as the traditional tools of B2B marketing and sales for identifying customer needs and pricing offers to generate demand, often by replacing activities which a business customer may previously have managed in-house.

There are three basic ways in which this may occur:

- Own use - a company does not actually share any data but uses the insights to refine and strengthen its own commercial activities, e.g. targeting specific customers, adjusting prices, refining the commercial offer etc.;
- Trading in – the actual sale and provision of data, insights or aggregated information, which may be a one-off snapshot or a continually updated stream; and
- Trading on – a company uses the insights to sell specific services, e.g. targeted advertising, without selling the underlying data itself.

The table below shows examples of these types of data use.

Table 1

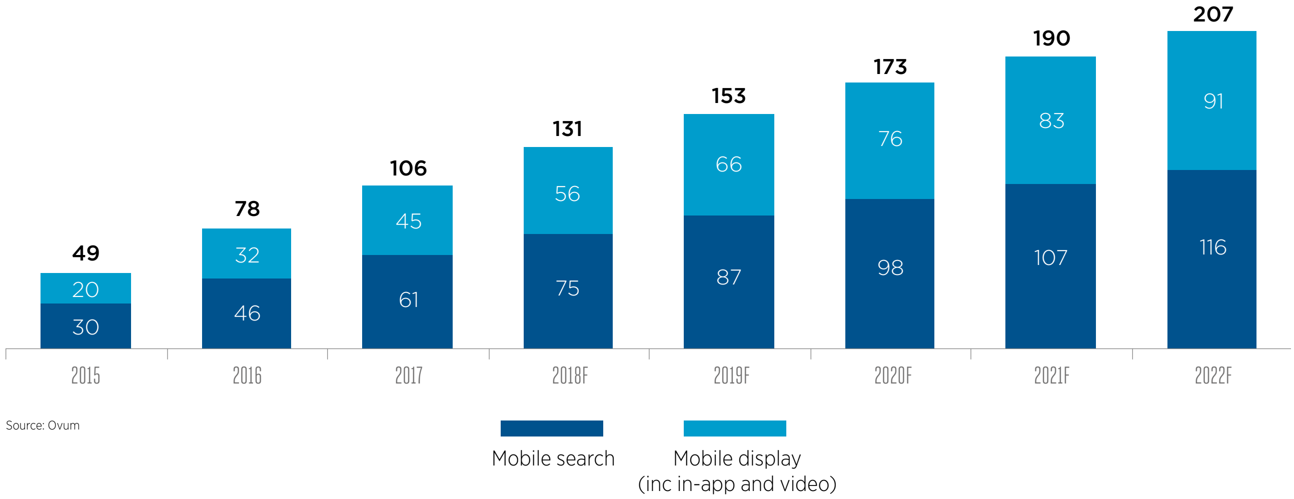
Types of data uses and trades

Exchange type	Examples of specific uses
Own use (Non-traded)	*Revenue optimization* – insights from data can be used in a multitude of ways to drive growth including the ability to improve product offerings, enter new markets, change distribution channel strategy and target consumer segments. *Cost optimization* – the use of IOT is revolutionizing the understanding that businesses have of their own operations. *Product creation* – development of services based on artificial intelligence, with the need for data to train any machine learning capabilities.
Trading in	*Market intelligence* – data relevant to a company's markets including the external environment. This can be used for decision-making in determining a strategy. Very few firms can generate sufficient data themselves. There are however a number of data aggregators who operate in this area who sell the data directly to businesses and other organisations.
Trading on	*Marketing and advertising* – data insights based on customer transactional and behavioural data from multiple sources such as social media profiles, demographic information, online browsing history and previous purchases give marketers and advertisers a head start in their ability to reach the right audience at the right time. Conversion rates indicate the value of the data. *Market-making* – this is the growing business model adopted by many internet businesses who match the needs of buyers and sellers through a platform. AirBnB, Uber and Tinder are all good examples. Scale is very important as the liquidity of the market created by the platform is a key feature demanded by users and in most cases there are strong direct and indirect network effects.

Online advertising has been a major growth area which leverages the data collected about individual users, both directly collected and observed by analysing behaviours that include interests, location etc, depending on the website and app being used. Companies that hold this data are then able to monetise this by serving adverts to the users if they meet criteria defined by paying advertiser. Figure 6 shows the rapid growth in the online advertising market worldwide and also that future growth is expected to come from mobile advertising, through a combination of internet browsing, search services and the myriad of apps that run on smartphones and include advertising.

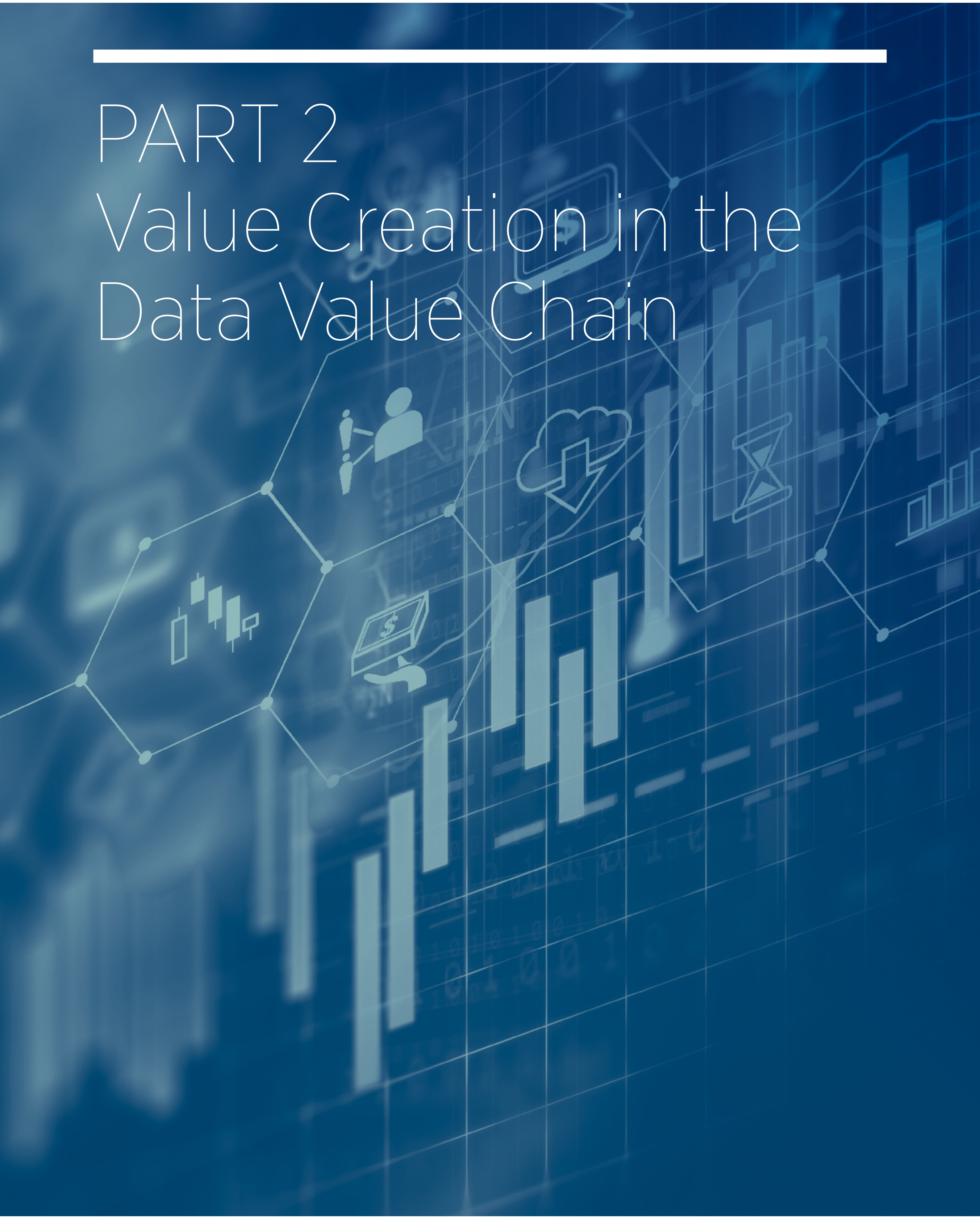
Figure 6

Worldwide Online Advertising market ($bn)



Source: Ovum

Many online adverts are tailored in some way to the user viewing them, whether it is by location, previous viewing history, previous searches, or more detailed knowledge of interests, social network and possible needs. It is the data collected through many sources and interactions, collated, mined and analysed that is used to link advertisers to users and potential customers.



PART 2
Value Creation in the Data Value Chain

Analysis of Value Creation at Each Step

1. Generation

Value Creation

Companies can create value at the data generation stage in two ways. First through their ability to capture and record the source data, at scale, in a useful format and economic manner. On the supply side, data is generally becoming cheaper to collect. For example, the average cost of an industrial sensor in 2004 was $1.30 and by 2020 it is expected to be $0.38[21]. However, this level of commoditisation may be hindered where access to data is limited by physical, economic or legal constraints. Mapping information and satellite imagery, for example, can be extremely resource intensive to collect so that once the map is created it becomes a very valuable resource. Exclusive proprietary data can also hold high value but this is hard to protect if the data can be captured or replicated in other ways. A store may believe that its sales figures are confidential, but if many shoppers share their transactions and observations on social media and CCTV reveals how many cars are parked outside, it becomes possible to predict with a high degree of accuracy whether the store outperformed its rivals or not over the course of a weekend. When data is generated and captured in such ways that are not easy to replicate then intrinsic value is created due to scarcity and

At this stage of the value chain, players can gain a competitive advantage at the generation stage through the scale and exclusivity of the data they capture. This can relate to the scope and comprehensiveness of the dataset about an individual or a transaction or an organisation, for instance the more detailed the information about an individual, the greater the potential value for advertisers. It can also relate to the scale of data held about a population. Having detailed information about 90% of potential customers or voters is much more valuable in many contexts than the more traditional approach of sampling and then extrapolating and generalising insights based on less than 10% of the population.

The breadth of services available at a zero price point via the internet shows the eagerness of companies to develop services that engage the widest set of consumers and persuade them to share data with the service, which in most cases is then used to target advertising[22]. The benefit of attracting more users generates many positive impacts in the form of returns from scale and scope, and the cross-group benefits of a larger user community, all of which increases the utility and value of the service. It is interesting to note that all the major search and social media services, which have the strongest network effects, do not even offer users an option to pay for the service in return for not sharing their data or receiving advertising. This differs from most ‘freemium’ type services that are selling a product or service (i.e. music streaming, data storage, etc.) and need to generate an income from this in the first instance, for which data collected from users is a positive externality but only the by-product of their main activity.

As the issue of permission to use data (‘consent’), and for what purpose, rises in prominence among the general public, companies that are able to gain trust from the data subject will be able to add more value to the collection of data. This would be particularly true if they could use this trust to gain more data points and therefore better insights (and not just collecting a higher raw volume of data). A company called People.io has developed a service which enables users to license the use of their data in return for being compensated with rewards or discounts for products and services. The selling point for advertisers is that they have explicit consent to send advertising and promotional offers to the user with increased relevance of those messages, and a workaround to the rise of ad-blocking. In this way People.io is enabling users to realise the value of providing consent to the use of personal data.

21. Ennomotive (2017) – Industrial IoT is booming thanks to a drop in Sensor Prices, https://www.ennomotive.com/industrial-iot-sensor-prices/ Accessed 23rd October 2017

22. As described in the Internet Value Chain (GSMA report, The Internet Value Chain: A study on the economics of the internet, May 2016), another rationale for the free strategy is to upsell customers on a premium version of the service, but advertising via data capture is still a major source of revenue

Trust

Trust ia a key requirement for platforms is to establish a trusting relationship between parties. For Facebook this is the trust it garners between both its users and advertisers. If trust is eroded then its raw material is threatened. With the rise of fake news leading to diminishing confidence and trust in its users it had to act quickly to build in tools to allow issues to be reported. The recent disclosures about how third-parties have harvested user data in the past also clearly have an impact on user trust. Facebook has also had issues with advertiser trust. It operates as a closed-loop system keeping its data in a 'walled garden'. However, there have been a number of disclosures recently that have cast doubts on Facebook's honesty toward ad buyers. For example, in September 2017 the company admitted that it had overestimated for years how much time on average viewers spent watching video ads on the site.

Coyle (2016) raised the question as to whether platforms can in effect act as a self-regulatory organisation due to their need to establish trust. This idea could be further extended to argue that the open review and feedback mechanisms of platforms could in fact be more effective than top-down mandates from integrated companies or institutions[23]. For example, in the accommodation sector, large global hotel brands, such as Sheraton and Hilton, build trust through defining extensive operating procedures for every detail (right down to the brand of soap provided in bathrooms) and auditing franchisees to ensure they fulfil the brand promise. Platforms such as AirBnB ensure quality and trust in a different way through two-way feedback ratings, which guests and hosts rely on for future bookings.

Market Features

There is a highly competitive market for data-generating applications and devices in many areas and any company wanting to collect new data is likely to be able find a supplier of the necessary hardware, software and if needed a service provider to collect data. Any organisation seeking to collect and use its own internal data is therefore served well. However, organisations which require unfiltered data from third parties – be this individual, machine or system – must find a viable route to data. For individual data, gaining access to the individual and gaining their trust and consent to access and use the required data can be much more difficult.

In traditional market-based activities, one expects normal commercial incentives to apply, such as a fee being paid to the individual in return for the use of their data and consent. The ubiquity of large, free-to-the-user platform-based services does lead however to different models. Users wishing to use a popular service such as a social media platform have few choices if they want to connect with other users already that service. Consequently, they have no ability to negotiate the terms on which they use such a service. It is interesting to note that no major social media service has a 'paid for' option that collects less user data and prohibits advertising.

In the industrial sector, the provision of real-time data capture may be offered as part of a lifetime maintenance contract for equipment. This may be efficient and valuable to the equipment user but makes it more difficult for alternative providers of maintenance services to compete without them being able to access the same performance data. System-generated data is often proprietary to the operator of that system, although there may be rival systems (e.g. for share price and trading data). Therefore, the collection of valuable data is often hard to separate from other steps in the value chain, which is one reason why it is impossible to quantify in financial terms the value of data as it is generated.

23. Though public institutions will always have a role in defining and enforcing minimum acceptable standards, e.g. trading standards, product safety and quality, food and hygiene regulations etc.

2. Collection

Value Creation

The ability to connect and transmit data between devices and storage locations is an essential function in the operation of the data value chain. There is a wide choice of options to achieve this, with a competitive range of technical features and associated prices. Key value drivers include network reach, reliability, security and performance. Depending on the application need, a service provider may choose to pay more for additional services or a better performance which enhance the value at the collection stage. For example, the UK government has commissioned a dedicated radio network to connect smart utility meters, rather than using existing networks, to ensure they can specify precisely the security and coverage parameters and control the service technology evolution over the lifetime of a meter. Other 'connected' type services place a high degree of importance on the ability to collect and share data in a real time and in a reliable manner. For example, there is currently a debate on the best way to do this for autonomous cars so that they can share data between each other about speeds, braking, traffic conditions etc., with the competing needs to share some data in milliseconds locally, but others over larger areas, which implies different technical requirements.

As more data is collected at more frequent intervals, even via continuous streaming, the importance of being able to validate, store and organise these large volumes increases. Schroeder (2016) in his research interviewed data experts who stated that at least 75% of data scientists' time is spent cleaning and preparing data for analysis with the remaining time doing the actual data science. The cost of this activity reflects the value that is created by curating a robust and reliable dataset to be used in other steps, including analytics.

Place ID is an example which demonstrates the relative value that can be created at the collection stage. The company uses a specialist proprietary technology called 'Darwin' to ensure data accuracy. This technology uses several methods to ensure dependability:

- Data spike detection – removing data if there are an abnormally high frequency of location signals for a particular location
- Data pattern analysis – removing 'artificial' data, detected through identification of odd patterns like perfectly shaped lines or boxes
- Data "noise" algorithms – assessing reliability of data where there are no obvious inaccuracies through using algorithms to detect low resolution or "noise" in subtler ways

Skyscanner is a service which collects airline prices and allows users to search and compare them in one place. Its ability to gather a multitude of fares from the broadest range of airlines is the unique feature which attracts users to its website. It acquires the fare data in two ways. The first is via an API set up with the airlines. The second is using web crawlers which run an automated script to extract price information. As Skyscanner has grown and evolved into the default market leader for flight price comparison, the first method of data extraction has become the norm since the airlines want to make sure their prices are easily available. Previously some would have tried to thwart the web crawlers in the hope of users coming directly to their own site.

Collection is important as a preparatory event to collate data and ensure quality and accuracy at the analytical stage. It is however difficult to build a distinctive position at this stage of the value chain. Standard network connectivity will suffice for many uses and can be taken as the norm in many economies, along with the interoperability of network infrastructures, most notably the Internet. However, it is important to recognise that this is not true everywhere, so that reach into remote areas can be a source of enhanced value creation. Equally, as requirements for the speed, reach, and QoS differentiation of data collection increase for IoT applications, there will be opportunity to provide value by serving different requirements.

Ten years ago, even the basic task of saving and storing data was a complex activity. It required a degree of manual intervention and significant proportions of data centre space capacity given over to specialist storage hardware (consisting of instant access arrays down to magnetic tape back-up and archiving systems), with different levels of availability and resilience, taking up a lot of physical space, and often managed in-house at high cost. Today, companies of any size have access to almost limitless storage capacity on a 'virtual' basis without the need for major upfront investment that would have been a barrier in the past. The increasing processing power and storage capacity available at ever declining prices have arguably created huge value for almost all players in the data value chain. There is a similar trend on the processing infrastructure side. What was previously an expensive and specialist activity for companies with high upfront costs, is now much more scalable and accessible for companies of any size. Even small companies can access powerful computing power on a pay-as-you-go rental basis.

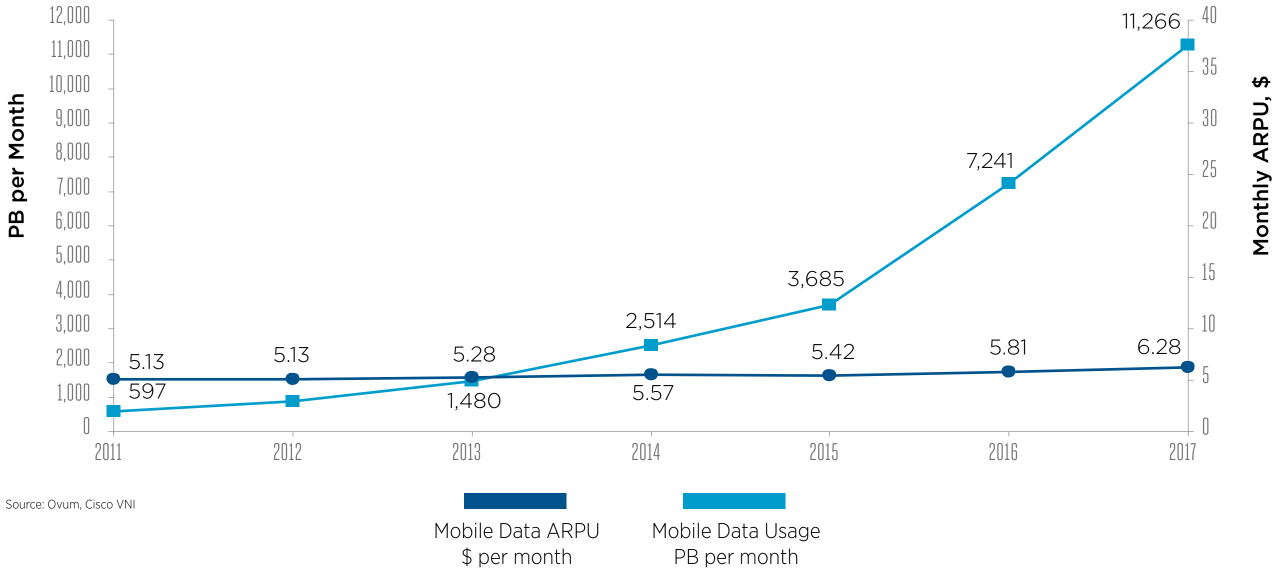
Market Features

There is a highly competitive market for the provision of data networks based on open standards in almost all countries. A clear driver of growth for many data-centric businesses is the increased access to ubiquitous fast broadband and the exponential growth of smartphones and other devices which allows data to be collected in most places at very low cost. Where more specialist needs exist and proprietary closed network infrastructures are needed, organisations have proven willing to pay for these and can choose from an array of technical and commercial options to do so, such as the smart meter networks, traffic monitoring services and industrial networks. The collection market seems set to continue growing at pace in volume terms but will only capture a modest share of the total value chain in the future given the competition and substitutes for the different elements required. As figure 7 shows, in recent years, telecom network companies (both fixed and mobile) have seen exponential growth in the amount of data they carry over their networks. However, the competition between providers has resulted in the average revenue per user (ARPU, a standard industry metric) growing only slightly on a global basis and actually declining for many years in more developed markets. The only growth comes from adding new users (and therefore more data traffic). Further, there has been increased growth in the use of alternate networks, most smartphone and mobile devices able to access Wi-Fi services when available as an alternative. As a result, the available infrastructure will only develop if policymakers ensure that conditions for further investment in networks are predictable and favourable.

39. GSMA Mobile Privacy Principles (2016), see: http://www.gsma.com/publicpolicy/mobile-privacy-principles
40. http://www.gsma.com/publicpolicy/mobile-privacy-principles

Figure 7

Mobile Data usage and Monthly ARPU (Average Revenue per User)



Source: Ovum, Cisco VNI

The market for the provision of storage has been competitive in terms of price evolution to date, with the unit price of storage falling continuously each year as operators build increased scale and the leading global players compete to win market share and grow the market by competing with traditional offline storage solutions. Prices of cloud storage have fallen from around $0.15/ GB per month in 2010 to just $0.02/ GB per month in 2017[24]. However, with the race to compete on price in a business with large scale effects, there has been a gradual concentration of suppliers, since only the largest and most efficient can compete: Google Cloud, Amazon Web Services and Microsoft Azure are now the major suppliers in the global storage market. Amazon Web Services is the biggest player in cloud storage and the fastest growing part of the company as well as its most profitable division[25]. To date the storage market has served customers extremely well but should be an area to be monitored in future as the potential for scale players to leverage their scale into other segments of the value chain could grow. It is no surprise that as a result, the cloud infrastructure market which was worth $12bn in 2014 is forecast to be worth $70bn by 2020[26].

Although the benefits in terms of costs are clear, there are some restrictions which prevent companies making full use of such scale. Many countries have data localisation requirements that require companies in certain sectors, e.g. banking and telecoms, to store and process data within the country. However, this results in sub-scale operations and higher costs for businesses that are not able to access the latest technologies or reap the benefits of scale in an activity that has large scale economies. Operator with networks in multiple countries and economic zones, sometimes up to 15, have to deal with complications and sometimes overlapping requirements relating to cross border data transfer and are prevented from building the scale that integrating these networks would bring, e.g. through shared data centres and common platforms. The data storage is also arguably less secure since it is primarily dependent on the quality of the infrastructure (including physical building, redundancy, back-up and disaster recovery facilities) rather than the geographic location of the storage.

24. https://www.backblaze.com/blog/cost-data-of-transfer-cloud-storage/
25. https://www.theverge.com/2016/10/27/13441508/amazon-q3-2016-earnings-report-aws-profit-miss Accessed 17 November 2017
26. Gartner Cloud Services report, 2Q17

3. Analysis

Value Creation

Analytics is viewed by data economists as the stage which has the potential for the most value to be created, since it is the stage with the most potential for companies to innovate, develop and deploy their intellectual property and by so doing, build up more attractive, defensible and higher margin businesses. The ability to generate 'self-learning' algorithms using artificial intelligence is a means to accelerate the development of such businesses (although there will also be losers who invest in such development but the resulting software proves to be fruitless.) It is not just the analytical tools which capture more value. The outputs of the stage such as business insights and presentations are much more actionable and more likely to be unique outcomes of proprietary algorithms (as well as perhaps exclusive data inputs) and are therefore considered to be worth more than the preceding stages.

As with any business linked to innovation and the development of intellectual property, the key success factor is access to talent, even above technology, which is generally commercially available to all for a price (as demonstrated in the section above on cloud processing services). Some academics such as Lerner (2014) even argue that at this stage data itself is not the most important factor.

Numerous businesses are capitalising on the attractive margins offered in the analytics space to offer additional insights to companies. One example of this is reputation management. Social media data is increasingly being used to provide companies with quicker, easier and less intrusive ways to gauge customer perceptions of their brands. Rather than scoring postings as broadly positive or negative or looking at well-known rating sites, businesses in this area use learnings to make conclusions from unstructured comment data on broader social media platforms to do with emotions such as disgust and anger. At present the broader technique of sentiment analysis - using natural language programming and other textual analysis methods to assign positive or negative sentiment scores, typically to posts on social media and forums - operates only at around 70% accuracy (though some providers have recently claimed rates of 80%). This is due to obstacles including sarcasm and irony, homonyms and an as yet unresolved inability to reliably detect the subject of sentiment in threaded conversations. As technological understanding evolves via machine learning, the reliability of such insights will increase.

While there is room for a lot of innovation in analytics services, clearly the raw input data is an important component. Unlike the three examples above, a new entrant with innovative software algorithms but no access to relevant data would struggle to monetise this directly and might prefer to sell its software (or the entire company) to companies holding the data. An example of this is Utiligroup, a company which coded the connections and data flows that are needed to set-up as a utility company. The service connects the various feeds of energy data together in a compliant way and processes them so that companies using the tool (in this case energy retailers) can comply with billing legislation. They sell this on a 'software-as-a-service' basis where a client pays them based on the number of connections processed but the company does not take control of the actual data passing through the system, which remains with the utility which owns the customer relationship.

Market Features

In general, the market for analytics functions effectively with many providers and a constant flow of innovation in tools and applications. Insights are generated by companies using their own innovations or procuring them from vendors. The ready access to computing power and storage discussed in the prior section removes one potential scale barrier that may have existed previously and so most companies can innovate and execute according to their abilities. Access to talent and intellectual property becomes the key differentiators as in many other knowledge-based sectors.

However, access to the necessary data is an area of potential concern and the ability of companies to control the source of the data needed could become a potential constraint that restricts others carrying out data mining and analytics. While talent and technology are the key differentiators needed to develop innovative new analytic approaches, algorithms and insights, these are still of limed value without the 'oxygen' of data to analyse. As mentioned earlier, some companies overcome this constraint by either selling their capability as a service or even selling the company and its intellectual property assets to a company that leverage them. The impact of having access to data, or not, is likely to be become even more important in areas using Artificial Intelligence ('AI') approaches. In the hypothetical case of two companies starting with the same software, the company able to access and process greater volumes of data (assuming quality is comparable) will develop better insights than one with less data to process since the refinements and advancements effectively accelerate with 'experience' (or in the machine case, increased processing cycles).

This ability to engage in data activities across multiple segments can lead to greater value creation opportunities. Companies providing multiple services, e.g. search, streaming, mapping, social media services etc., can build up a more complete picture of individuals, how they use different services, a more detailed list of their interests etc. The comprehensiveness of analytic insight about a specific target could become a barrier to entry, as it would be difficult to replicate. There is of course nothing inherently wrong in the economic sense (the implications for privacy may be of greater concern) and this is reminiscent of discussions about competition between hypermarkets and high street specialty retailers. However, when combined with the vertical integration discussed elsewhere, whereby a company generates, collects, analyses and trades the data in a closed loop model, the potential for market power imbalances will be ever-present as the data value chain evolves.

4. Exchange

Value Creation

Each way of using the data end product has its own value creation profile.

a. Own Use

In the case of own-use of the data and insights, maximising the value here really relies on a company's ability to develop commercially attractive services based on the insights they have developed, or to use the insights effectively to refine their existing business and marketing efforts. These could be insights in how to better target offers and carry out segmentation, or operational insights that improve the reliability and usability of products or the efficiency of production facilities. In the retail or media segments, providers use insights on a customer's purchasing behaviour to make suggestions for additional or subsequent purchases. Retailers with rich reward programmes can build up detailed databases of typical buying behaviours and correlations and so are able to anticipate future needs. In the B2B area, it is standard now for manufacturers of sophisticated machinery, from aircraft engines to manufacturing robots, to have direct connections to the units they have sold so they can track their performance in the field and then use data reports to predict component failures. This reduces downtime and maintenance costs for their customers as well as improving their own R&D for future product launches/software releases.

Internal data, combined with market information such as published prices points, can also be used by companies to refine pricing models for their products and services as they understand better the relationship between price points, consumer behaviour and competitor reactions. Increasingly companies are using computer algorithms to take into account many variables and set their own prices. Some even go so far as to use inputs about the potential buyer, such as location (affluent or lower income area), access device (expensive or not) and previous buying habits and history (tastes and price sensitivity) to adjust prices that individual consumers see in order to charge a higher price to less price sensitive customers, an effect known as dynamic pricing (see call-out box). While such algorithms can increase efficiency for the seller, concerns have been raised[27] about the potential for such algorithms to be used by companies to collude on price setting without what would previously have required human interaction and agreement (however covert).

Dynamic pricing

The use of dynamic pricing allows businesses to maximise return on a product using data to inform more detailed price differentiation plans where the price a potential customer is presented with is informed inputs including geographic location, previous buying history, even the device they may be browsing on. When a platform wants to encourage more users or drive revenue growth they can lower prices and use targeted strategies towards consumers they want to attract. Amazon alters its prices more than 2.5 million times a day and changes the price of around 20% of its total inventory. It uses techniques such as geo-targeting. For example, if a consumer is thought to be near a brick-and-mortar outlet selling the same product a discounted price will be shown. The travel site Orbitz made headlines when it was revealed to have calculated that Apple Mac users were prepared to pay 20-30% more on hotels than users of other desktop brands, and to have adjusted pricing accordingly. Clearly the concept of price differentiation is not new but data analytics enables retailers and businesses to do it at a much more granular level than previously.

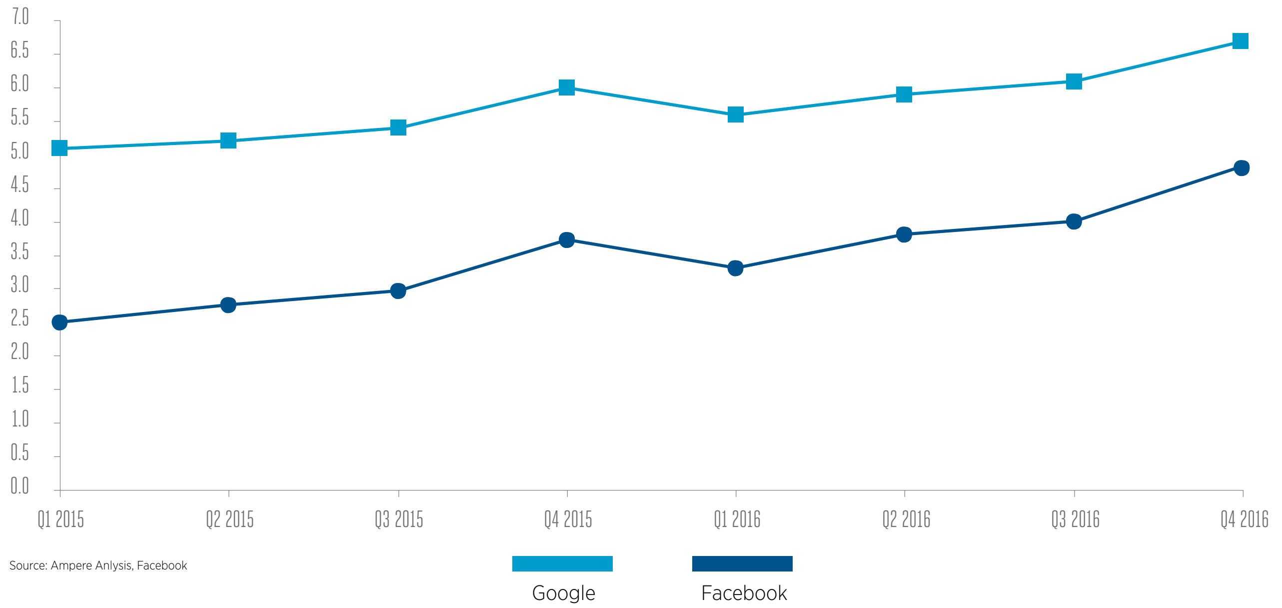
27. OECD (2017), Algorithms and Collusion: Competition Policy in the Digital Age

b. Trading in data
In the sale of data, value is derived from the quality, volume and uniqueness of the data service provided. Examples include the provision of energy trading streams, share price information and proprietary market research services. Data brokers facilitate and sell data exchange services. For example, Oracle is developing an exchange for data assets. This allows customers to trade personal data, combine them with sets provided by Oracle and extract insights. This is offered in the safe environment of Oracle's computing cloud where it can make sure that information is held securely and not misused. A key component of realising the value of data at this stage is the supplier's reputation for quality and integrity since all buyers will want to know the accuracy and provenance of the data. It is also an important part of the packaging step that companies develop the necessary data formats and APIs to transmit and share data. Making this as frictionless as possible makes the commercial proposition more attractive to potential buyers.

c. Trading On Insights
In this case, companies with packaged insights and information sell services to other companies, without actually selling the underlying data. The value comes from the breadth and depth of insights a company can offer to buyers. The most common examples are advertising based services. In these cases, advertisers can specify a multitude of metrics that describe their target audience and then the service companies (such as search or social media firms) will place the advertisements in front of those users when using the service. There is clear value for advertisers in being able to micro-segment the audience base including using socio-demographic criteria, age, location (of residence or even at a particular moment), interests or even links to other people in the case of social media networks. It is therefore no surprise that the large global players excel in this area, with 63.1% of US digital advertising investment being captured by Facebook and Google[28]. In fact, as figure 8 shows, Google and Facebook have been able to increase the revenue they generate from each active user, demonstrating the steady increase in value of online advertising. Other services may be able to target particular niche audiences due to the nature of their services, for example a cinema booking app or a cycling app.

Figure 8

Quarterly revenue per active user, worldwide ($/quarter)



Source: Ampere Anlysis, Facebook

28. Taken from https://www.emarketer.com/Article/Google-Facebook-Tighten-Grip-on-US-Digital-Ad-Market/1016494 Accessed 17 November 2017

In terms of the commercial sale of data, at least presently, there is not a strong trade or open market for raw data. Research commissioned by the European Commission as part of a study in the European data economy[29] found that 78% of companies surveyed generated and analysed data in-house (or by sub-contractors) and that only 4% of companies reported data trading among independent entities. Quite apart from the need to access and validate data (see above), there is a lack of data fungibility making it difficult for a price to be put on a specific set of raw data. Therefore, one-off sales are difficult (an exception may be the download of confidential data via hackers or whistle-blowers, but even in this case the data value is only fully realised after careful processing and analysis of the data). The importance of own use applications illustrates how often the actor which controls the data source is the best placed to realise the value of the data. It may purchase interim value chain activities such as storage or elements of analytics, but keep end-to-end ownership of the data and manage the exchange of the resulting outputs for commercial benefit in compliance with data protection regulations.

Where data can be aggregated and anonymised, it has more potential to be collected and sold and there are emerging marketplaces that facilitate the buying and selling of data. The leading examples of data broker services are for aggregated and anonymised data, as well as personal data, for marketing purposes, e.g. Acxiom and Datalogix, or very niche data such as new car registration information and energy trading statistics. Dawex is an example of a platform that seeks to match buyers and sellers of such data specific to industries and provides the facilities for companies to transact and control data trades in a secure way and that meets the many requirements on privacy and security. The aggregation of large volumes of data addresses quality concerns and anonymization can overcome the trust aspect. There are many examples where medical details, such as results of experimental diagnostic trials have been collected and used in this way. Sophia Genetics is a Swiss company which has developed an analytics platform to help hospitals analyse genome sequences whereby a user in the hospital shares the raw data and the company's artificial intelligence platform is able to carry out a cancer diagnosis and charges for this on a per patient basis. The more data hospitals share on an anonymous basis the better the AI algorithm becomes and further helps clinicians elsewhere. In this sense there is a true "exchange" taking place that enables more value to be created for end-users in a virtuous circle.

29. European Commission Staff Working Document on the free flow of data and emerging issues of the European data economy, January 2017

Market Features

The existence of such services indicates that generally the market for data exchange works according to normal commercial trading rules and principles and, in most cases, there are various competing companies offering similar services so buyers have a reasonable choice of competing alternatives. It is certainly observed that there is not a general, liquid market for data outputs. If there is a perceived problem with low levels of data trading this would appear to be primarily due to the practical difficulties of transacting, as discussed earlier, especially at the early stages. It is possible that as markets mature and technology advances such trading will become more feasible and the barriers to trade overcome. There are certainly such expectations among industry commentators and forecasters. The emergence of services such as Dawex are perhaps and early indicator of this, although it is too early in the development of these services to draw such conclusions.

While difficult to prove, it is likely that for many businesses a sizeable proportion of the data that they collect is of little or no value other than they form part of the overall data pool from which valuable insights can be derived. It is necessary to collect the vast set of data before a company can be in a position to determine which data is valuable and contributes insights and knowledge. For example, when analysing traffic patterns in a city it is likely that the information about routes taken during peak times is the most useful but it will be necessary to collect data over the full day to be able to determine when the peak times are and to then drill in to the patterns during those times. Another example would be the data which is collected to track user behaviour of many websites. Many websites use a variety of means to track how each user behaviours and uses the website, from the user's unique ID (often collected via the ubiquitous cookies to link to previous visits), the time of day, how long the user spends on the website and how they navigate through pages, the device they are using to access the website right down to how long it they spend on each page, whether they scroll to the bottom and how long they take to click on each link. All this observed data, collected for every single visitor quickly generates a large record of behaviours that the website owner can use to improve the experience for customers and also the efficiency of providing the information or service they intend, but it is only when they develop hypotheses and design queries to understand the pattern of user behaviour that the data has any use or value and that significant portions of the data are not subsequently used but this cannot of course be known at the time of collection. In this case, that data is of minimal value to any other users either, unless it could be combined with similar data from other websites and generalised learnings developed about how a population interact with certain types of websites and the insights codified in a way that could be used to inform better website design for others.

These points highlight that any assumption that all types of data are available freely or commercially is not true. There are clear benefits to being as close to the source of the data as possible for companies that have a business model that can capitalise on these attributes and add value through building instantaneous or exclusive access to data, and being able to determine the context and structure of the data. Companies naturally seek to organise themselves and their activities to maximise these benefits.

The one area of potential concern is linked to the conglomerate effect identified earlier. The ability of players to build substantial platforms or to build insights from across many service areas gives them a very strong position in the market to sell their services, as the dominance of Google and Facebook in the US online advertising market demonstrates. In fact, they are expected to increase their share of the global online advertising market to as much as 84% of global digital advertising spend (excluding China)[30], which raises concerns about an emerging duopoly. The network effects inherent in their business model and services – which can have a benefit to consumers – make this a difficult issue to address. A further concern is that closed loop models might restrict competition if they are built on data which is hard to replicate, but there are few obvious examples where data is totally exclusive to one provider.

An important point that should not be overlooked at this stage is that companies that have invested in collecting and processing data up to this point need to be able use it for normal commercial purposes, within the bounds of what is permitted by privacy laws. Any interventions that might try to set a predetermined price or require companies to share data publicly for free would affect the investment case for collecting data and innovating in how to use it. Around the world, there have been various governmental initiatives to make publicly funded data more freely available, such as census records, academic papers and research, etc., since this has already been paid for from public funds. However, these initiatives should not be extended to privately-held data. Where the data has been generated fairly but commercially, it is important to protect the providers' interests and stimulate innovation, provided they do not raise other competition concerns.

Types of business model

Within the framework and activities described above, there are several ways that companies are organising themselves to provide services and creating value. In a different value chain, say for example the extraction of iron ore to make steel products, companies would often specialise in one activity, e.g. mining, processing, sheet metal production, and then sell their outputs to the next player. However, in the data value chain it is immediately evident that many players are active in multiple steps of the value chain. Where companies do specialise in specific tasks, from storage, Software-as-a-service (SaaS), specific tracking and data analysis, this is on a sub-contracted basis where they provide the service to the company controlling the data but do not have rights on the data and cannot use either the data or the insights developed for any other purpose. EU Commission research into the emerging issues of the European Data Economy using a survey of businesses found that while a majority of companies sub-contracted data analytics services, this was *"only for the specific purpose and within the limits of the contract. Further re-use of the underlying data is typically not allowed."*[31]. There are several reasons for this that stem from the characteristics of data.

Firstly, due to its unique characteristics it is quite hard to value and therefore trade data, especially at the early steps in the value chain. There is no consensus or emerging models on how to price data as a tradable asset in the market pricing sense (as distinct from services and revenues that the data may support). Even from a legal point of view, ensuring enforceability of a contract trading data and information (and making sure it is not copied and passed on, or used for other purposes than those agreed) is not an easy task.

Due to the uncertain or unknowable quality of data when trading it, companies that use the data which they collect themselves are in a better position to apply any feedback to further increase the value of the data

30. GroupM – Global Ad Investment report; https://www.groupm.com/news/groupm-global-ad-investment-will-grow-43-in-2018-six-countries-to-drive-68-of-incremental-investment
31. European Commission Staff Working Document on the free flow of data and emerging issues of the European data economy, January 2017, page 15

at any stage in the value chain e.g. by more timely collection, better storage and indexing, or formatting to support later analytics. If such innovation took place as a set of transactions between independent players, it would be harder for such improvements to be implemented and rewarded. The seller would need to prove that their data is of higher quality and the buyer would have to spend considerable expense to validate this and discern between different sources, especially in a rapidly chaining and innovating environment. By carrying out each of these activities in-house, companies can minimise the transaction costs and focus on the process of data collection through to analysis. Since such costs are significant, there is a strong driver for integrating these activities within a single entity.

There are also the requirements on data collectors to protect the privacy and use of the data they collect. As a result, any company with personal data is likely to find it easier to process the data itself than it is to sell the data to another player, even if they have better analytics capabilities. In fact, they are much more likely to buy in the capabilities than to sell the data, and so in most cases will retain responsibility for the data directly.

Many companies have a 'single-service' approach and use data to enhance and develop their service. The examples of businesses that use data about their users to improve and better target their services fall into this category, as would information providers who collect data, process it and sell insights and services on to third parties. In the case of trading-on data, companies are providing a service, which is based on the underlying data but the data is the product being sold. Even in an apparently clear-cut example of information services for share prices, the share price information feed is actually the output of a series of analytic steps that derive the price from a set of discrete transactions using an algorithm.

Platforms

A common feature of many internet and data-driven businesses is that they are platforms. There is no clear definition of what constitutes a platform, a fact recognised by Coyle[32], but there is a reasonable consensus that platforms are digital infrastructures that enable two or more groups to interact. They therefore act as intermediaries that bring together different players in the data value chain: customers, advertisers, service providers, producers, suppliers, and physical

32. Coyle, Diane, Making the Most of Platforms: A Policy Research Agenda (October 17, 2016). Available at SSRN: https://ssrn.com/abstract=2857188

objects. It is also true of most platforms that the more people that use them (from whichever user group) then the greater the utility of the platform and hence their value creation potential. In such cases, there are clear consumer benefits but a clearly market leading platform has very strong market position. A good example of a platform that delivers a clear consumer benefit by bringing the largest group of potential buyers and sellers together is eBay, the online auction and direct selling website. By being a single destination for the buying and selling of a broad range of products from the trivial to items such as cars and works of art, it creates a platform where buyers can find a broad range of products they may be interested in and sellers can access a large audience of potential buyers. In the course of running its services, eBay collects data on buyer searches, transaction histories and also the rating system that underpins the trust that is needed to enable strangers to transact over the internet. As well as the direct network effects which favour a large auction platform, the data behind it provides additional indirect effects which further strengthen the service, and with it, the position of eBay as the leading auction provider.

There are distinct types of platform, depending on their core revenue-generating business, including the following examples:

- Advertising platforms – these are characteristic of many social media players. They use their detailed knowledge of users to enable advertisers to target users that fit selected criteria e.g. Facebook
- Cloud storage and processing platforms – these organisations operate as Software-as-a-Service (SaaS) or Platform-as-a-Service (PaaS) and therefore link business with third party developers through an API e.g. Salesforce
- Product platforms – turn traditional good into a service and collect rent e.g. Spotify
- Lean platforms – these enable asset sharing between asset owners and users without the platform taking on a responsibility for the asset themselves e.g. Uber, AirBnB, ride sharing apps

Platforms can also operate in two ways, either as an open system, with an ecosystem of partners involved in providing the overall service (although the data itself may only be accessible to the platform provider), or as a closed system with close integration for data collection to output.

Looking forward, it is possible that as business models mature, technology advances and trust is built, new ways of working may emerge that overcome the challenges to data trading and transfer and so vertical integration reduces. Companies that specialise in one activity are able to excel without needing to invest in and address the full value-chain of activities. In Europe, the PSD2 regulations coming into force in 2018 will create the possibility for consumers to request their banks or credit-card companies to transfer data on their financial transactions to a third party. This can then offer additional services, including amalgamating transaction from different financial providers and performing analytics on a consumer total spend, and potentially make recommendations based on this analysis. In this case, the data collection and aggregation carried out by the bank and credit-card companies is separated from the analytics service provided by a separate company.

Case Example

Closed System

Duo Fertility is a company which manufacturers a device that supports natural conception. A patient secures a sensor to her body to record sleeping patterns, body temperature and other data, all of which is stored. The data is then interpreted by DuoFertility's proprietary analytic software to calculate levels of fertility. For couples hoping to conceive this provides very useful data which is viewed via web-based software. DuoFertility manages the entire data value chain including the sensor device, the data processing and feeding the results to the user. The proprietary devices for data collection are linked to a closed analytics platform with no ability for a consumer to transfer their data elsewhere. In this case, the closed loop aspect is necessary to ensure privacy and security given the nature of the service but this also protects the company's innovations.

Case Example

Open System

Strava is an online service that allows runners and cyclists to log and upload runs and rides and compare them with others, creating virtual leader boards. In this case Strava does not manufacture or sell devices, the data can be captured on an array of GPS enabled devices such as running watches, smart phones etc. Strava collects the data from many users into one place and provides the platform and analytics to compare and display the results. The ability of Strava to collect an individual's data and compile it with that of others to form a social network of athletes creates a network effect, driving use of the app. Strava utilizes the data in several ways. It develops its proprietary analytic capabilities for users and offers Premium services for paying users. It also trades in collected GPS location data with third party companies, although it courted controversy recently when it made public the aggregated and anonymised running and cycling maps of all its users, which inadvertently highlighted the location of secret military bases, shown as areas of high activity in otherwise deserted areas. Interestingly, Garmin, one of the leading manufacturers of the GPS devices launched its own logging platform but this has proven less successful since this excluded non-Garmin users (and also because their devices are compatible with Strava) for a service where number of users is a key component to success.

Multiple Services

A different form of integration occurs when companies expand and operate in adjacent or even unrelated areas, either launching new services themselves or acquiring other companies. Conglomerates such as GE, Samsung and Toshiba exist and the primary rationale is the ability to spread risk, finance costs and overhead costs such as branding. However, in data-driven businesses, in addition to these benefits, additional benefits accrue to companies that are able to play and compete successfully in multiple areas and combine data and insights across the combined entity. By operating in multiple areas, companies can collect data from more sources, perhaps serving the same users or customers, and then combine the data and insights to create additional value.

An example is Google's Customer Match service that allows an advertiser to upload a file of customer data (name, e-mail address, etc.) and then Google can match this to its own users of its search, YouTube, Shopping and G-mail services. The advertiser can then target adds to its existing customers across these services because Google users use the same log-in profile for all services. It is also a stronger proposition to advertisers who are able to target specific groups of customers across multiple touch-points in a joined-up way and confident that they are addressing the same people in a way that is not possible in the offline world. For example, an advertiser wanting to increase brand awareness among 18-24 year olds interested in football could use the Google service to place adverts in front of their targets when they search, watch content on YouTube or review e-mails, and all in a very targeted way. In the offline world, to have a similar impact the advertiser may need to place adverts inside a football stadium, during breaks in televised matches and in sports related print media, and at much higher cost, and not be sure that members of the target group are even viewing the adverts in the multiple locations.

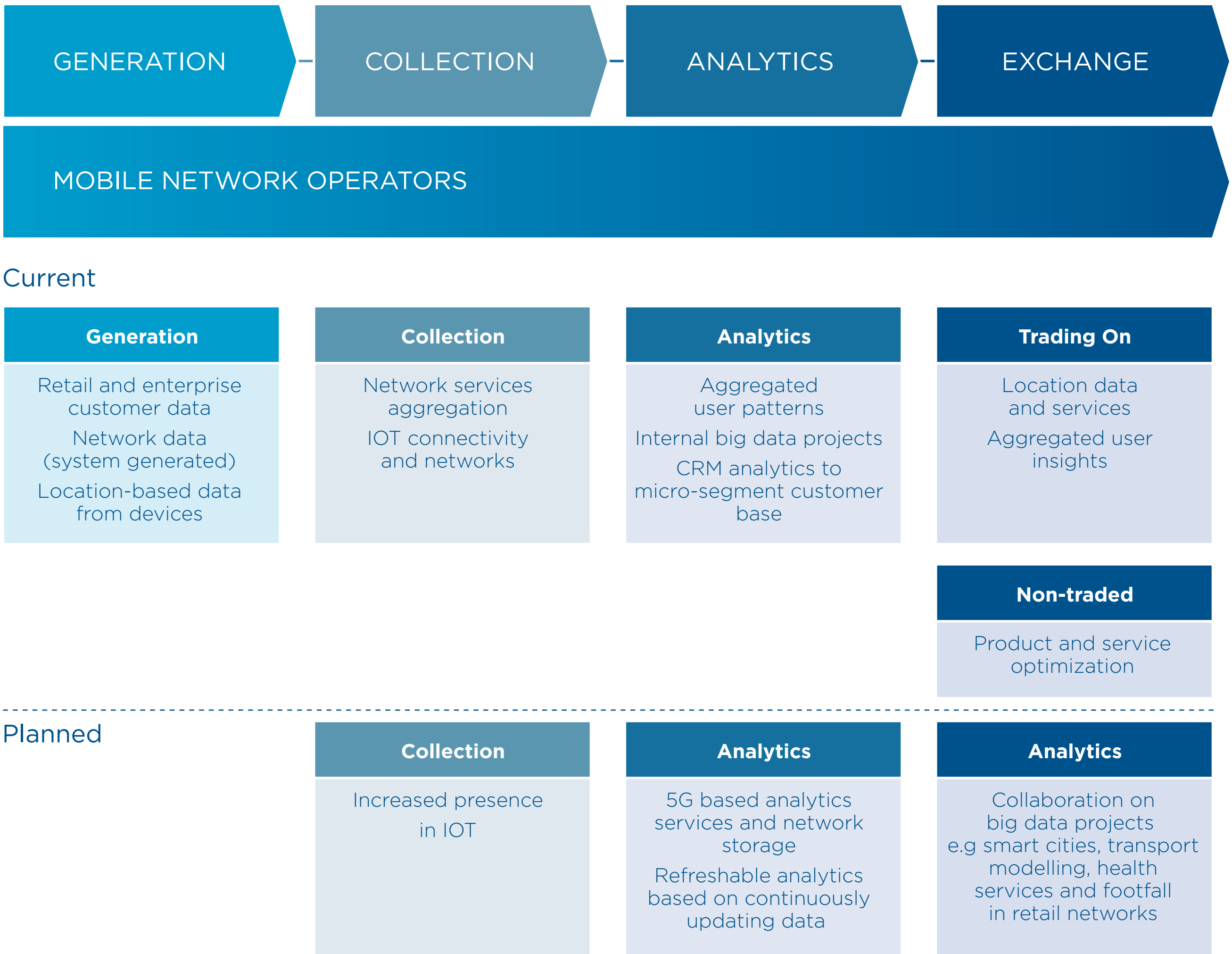
32. Coyle, Diane, Making the Most of Platforms: A Policy Research Agenda (October 17, 2016). Available at SSRN: https://ssrn.com/abstract=2857188

There have been examples of internet companies making acquisitions of companies providing services in other segments in order to be able to pool the data about users and so enhance the depth of data insight about the user base and so better target customers. Microsoft's acquisition of LinkedIn gives it unique insights into the trends, topics and the behaviours of networks of business people who are key customers of many of its products and services. Facebook's acquisition of WhatsApp generated considerable debate, and more recently concerns have emerged about the data WhatsApp is sharing with Facebook, such as user identity and phone numbers, that enable Facebook to match user accounts across the two platforms, indicating the potential value of merging the datasets to gain more insight on each user. In October 2017 the EU Data regulator group forced WhatsApp to pause the sharing of user data with its parent group without users' specific consent to link accounts and formed a task group to further investigate and resolve the data sharing issues.

Role of mobile operators

Figure 9

Overview of Mobile Network Operator activities in the Data Value Chain

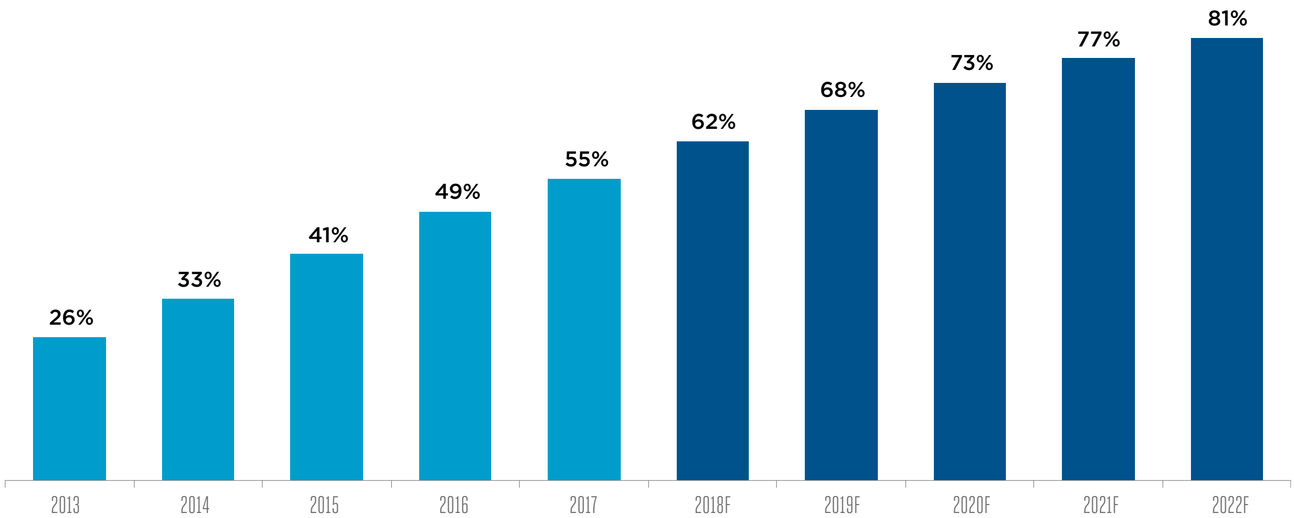


Source: A.T. Kearney interviews with GSMA operators

As figure 9 illustrates, mobile telecom operators are active in all parts of the data value chain but with a wide range in the extent of their involvement. As providers of data connectivity they are important players in the collection step, providing the coverage and bandwidth for all human and machine interactions. They are the primary means of connection for smartphones which are a major source of data. Figure 10 shows the rise in smartphone penetration worldwide, currently around 55% and expected to rise to 80% by 2022, which is enabling and driving this expansion.

Figure 10

Penetration of smartphones worldwide



Source: Ovum

However, it should be noted that in many cases they act as the conduit for data but will often not know the source or destination of encrypted data. Because of the use of VPNs, proxies and encryption, they generally will not even know which application on a phone the data is generating the data.

While running their networks, operators collect a certain amount of metadata, such as the location of a user through mobile radio cells (though not with the precision of the GPS function of a phone, which applications on a phone may be using), the amount of data being sent and received and the volume, length and source/destination of calls and messages sent. Mobile operators also collect personal data on their customers when they sign up to a contract as part of billing and identification checks, as with any other retailer or subscription-based service, which are subject to the relevant privacy and data protection laws. As part of running a complex infrastructure, operators also generate a large amount of operational data about traffic flows, peak loads, congestion spots etc, which they use to monitor and refine the overall performance of their network, and plan for investments in new roll-outs.

Potential Growth Areas

Many, but not all, mobile operators are investing in new data-driven services that go further than these core services, covering generation through to data analytics. The growth of Internet of Things applications and devices presents a large opportunity to connect a wider variety of devices. Some companies are focussing on connectivity solutions. Others are going further into device management solutions, either directly or as part of alliances. The development and deployment of 5G mobile networks in coming years will also create the opportunity for operators to increase the scope of their activities by building new services that require a high density of devices or specific low-latency networks for real-time interactions. Connected cars are another opportunity generating much interest at present, given the importance of mobile networks to enable these types of services.

Case Example

Telefonica Next

Telefonica Next is a subsidiary of Telefonica Deutschland which has been set up as a separate entity from the parent mobile network business to focus on data analytic and data-driven services. It offers a range of services including some which draw on the anonymised data from its network division, based on 48 million subscribers in Germany. Using this large dataset, it can offer services such as transport analytics for city authorities and transport companies. It has also developed a mobile advertising service that enables advertisers to target potential customers as they pass through specific locations such as airports or shopping malls.

Other related services including developing Internet of Things applications and a data anonymization platform. This service enables companies to anonymise large datasets in to a form which could then be shared and combined with other data sources to develop a range of insights about customer behaviours for example (but not at the individual level). The anonymization service removes personal identifiers and ensures compliance with data protection and security laws, of which Germany's are among the strictest in the world.

Case Example

Turkcell

Turkcell, the largest mobile operator in Turkey, has also developed a full digital suite of offers for their customers. In their case they have integrated data and analytics into platform type services such as messaging app, search, music streaming and e-commerce services including user authentication (based on Mobile Connect) and secure payment services.

Big Data for Social Good

The data generated and collected by Mobile network operators has the potential to be used for societal benefits by providing insights into the movements and actions of very large populations, given the majority of people in most countries now has a mobile device. As recognised by Anthony Townsend, senior research scientist at the Rudin Centre for Transportation Policy and Management at New York University, *"We've just deployed the best transportation sensing network in the history of mankind, completely by accident."*[33], referring to the growing prevalence of smartphones. Anonymised data from mobile networks (in the form of anonymous statistical information) has been used in cities such as Boston, Rio de Janeiro[34], Istanbul and other to help transport planning since the movement, origin and destination of journey's and travel versus dwell time can all be monitored. Big data analytics should continue to be used to identify patterns in the overall population that can inform demand for bus and transport routes, plan exit volumes from metro and subway stations and measure road traffic volumes (as well as destinations) with greater accuracy and insight than roadside counters.

In the UK, the government statistics agency, which is responsible for carrying out the national census every 10 years, is currently researching how it could use mobile phone data to more accurately measure things like population density, work commute flows[35]. Data from mobile networks also has potential to be used in other "smart city" applications. Singapore has been developing a range of smart services including real-time demand-driven public transport where shuttle buses run and reroute to respond to demand of commuters using an app, providing much more flexibility than predetermined routes and frequencies, as well as services that offer smart traffic management, cashless payments, health and flood warning systems. These use a range of sensors, interfaces and data inputs and depend on good integration with mobile networks. In a different context, mobile networks play an important role in disaster response efforts such as hurricanes or floods. Relief and emergency services depend on mobile networks to be able co-ordinate relief efforts but networks, where they are still functioning, can provide important and up to date data on where populations are located, for example in remote areas, so that help can be directed to them as quickly as possible[36].

33. MIT Technology Review, 18th November 2014
34. Analyzing cell phone location data for urban travel, Colak, Alexander, Alvim, Mehndiretta, Gonzalez, MIT & World Bank
35. Office for National Statistics, Research Outputs: Using mobile phone data to estimate commuting flows, www.ons.gov.uk
36. More details on such programmes can be found at https://www.gsma.com/betterfuture/bd4sg/

Barriers and Constraints

It is important that operators have a degree of freedom to use network data both at the aggregate level in the operational running and optimising of their network, and at a more individual customer level to design and promote different service offerings and price plans. Where it is the case, many operators are building teams specialising in big data and analytics to support this. However, in the cases where they develop other services that they may trade-on or trade-in, many operators consider themselves to be constrained in how they can analyse and trade this potentially valuable data. The situation varies from country-to-country, but a common situation is that an operator must secure explicit consent to use metadata and other information, or 'opt-in' consent before being able to push personalised services, while other non-operator providers of similar services, who for example may process GPS location data as part of their service, are not subject to the same strict requirements.

For example, developing business models centred on the analysis of movement patterns on a large scale and over a longer period of time, is an opportunity that mobile network operators in Europe will not be able to capitalise on. In Europe, telecom operators specifically are subject to the EU e-communications directive ("ePrivacy Directive"), which will be replaced by the ePrivacy Regulation currently being discussed in Brussels. It places an additional set of requirements on them which apply not just to telecom services but also when they are offering other digital and data-driven services and competing directly with other non-operator companies who are not subject to these requirements. For big data analysis, which needs a critical volume of data to get to reliable results, consent-only solutions will make it difficult to gain the critical mass necessary for robust analysis over time due to missing identifiers (momentary snapshots limit the potential insights and value)[37].

As another example, mobile advertising and the concept of pushing adverts to users based on their profile and location, e.g. special offers when walking past a particular store, is an opportunity that mobile network operators cannot capitalise on in all jurisdictions. This is again due to the requirements to obtain user opt-in and, more generally, a higher standard of compliance on use of personal data, whereas search and social media services have been able to generate large revenues from the same proposition and gain a unique market position.
Data localisation requirements present another challenge for operators in certain regions, with data protection restrictions preventing storage of data other than the country it was produced in. This can be especially problematic for operators operating in more than one country, and across continents where data protection regulations can vary significantly. The additional compliance demands drive up infrastructure costs and reduce economies of scale advantages that could be realized, limiting the benefits that can be passed on to customers. Operators, even in larger countries, are therefore not able to take full advantage of large cloud-based infrastructures that would reduce their costs for storing and processing data. Non-telco companies who are not covered by telecom specific regulations on the other hand, are able to do just that.

37. See also the GSMA's "The Proposed European ePrivacy Regulation Use cases for enabling privacy-protective innovative products and services, https://www.gsma.com/gsmaeurope/wp-content/themes/theme_europe/eu/GSMA_ETNO_ePR_Use_Cases_April_2018.pdf

PART 3 Issues and Policy Implications

Market issues

As described in prior sections, the data value chain is evolving as more sectors of the economy take advantage of the opportunities created by the availability of data and analytics. The value chain contains specific segments, such as storage, which have become mature and are broadly competitive. It also has segments where the market is less mature and where there are challenges in creating a smoothly functioning competitive environment, not least in determining market values for data as it moves through different activities and degrees of enrichment. Overall the rapid and sustained growth in data-driven businesses and services, for consumers and enterprises, is evidence of great innovation and it has driven significant efficiencies. Nonetheless the analysis of the market characteristics does identify some specific issues which exist or are emerging which require closer inspection, raise policy questions and may point to the need for policy interventions in future.

Direct and Indirect Network Effects

Many of the data business models discussed aspire to be platforms and therefore, since they are connecting groups of users, large platforms benefit from positive, self-reinforcing network effects. For example, Facebook will need to continue to innovate to attract and retain users and business partners but can do so with a high degree of confidence given the strong direct network effect which allows it to attract and retain its user base. The closed model whereby almost all interactions are via the Facebook app or website, further strengthen this hold. Such a model and the market dynamics give platforms greater power to set the prices and, importantly, the terms for users who may have little practical choice but to accept. Social media represents the best-known platform model, but this can apply to auction sites, or price comparison sites for a specific vertical. As well as attracting users and advertisers, or buyers and sellers, successful platforms often also create sustainable ecosystems where third-parties build new applications and functionality that serve many needs. Some are aimed at the users to increase engagement (and so potential for advertising), e.g. social games that go viral, others may address the needs to companies using the platform, e.g. tools to help them track social media interactions. All of these add to the attractiveness of the platforms but further multiply the 'winner-takes-all' network effects.

Indirect network effects support user growth and greater benefits, and enable businesses to grow but in many cases, there is also a 'tipping point' in terms of the commercial model. A good example of this are the price comparison websites that aggregate pricing data from providers of services such as travel, utilities and financial services. Initially the providers of travel or financial services may try to resist or block price comparison sites gathering their prices (a practice often called 'screen scraping' where an automated robot goes to the providers' websites to get prices in bulk by emulating user behaviour). The price comparators therefore develop more sophisticated, and costly, techniques to be able to gather the pricing data that they need to provide an effective comparison service. Once a service becomes very popular and large with many potential users however, then the primary service providers face considerable commercial pressure to co-operate with the price comparators in order to ensure their services are fully visible to potential customers. The next tipping point may see the platform reach the point where companies pay a commission or advertising fee to appear on the price comparison service.

These effects are often further reinforced in data-driven businesses through data feedback loops. For example, the more consumers use a search tool (either general internet search or a retail product) and a provider is able to build up profiles of how users interact with the results, the better they are able to refine the algorithms and make future search results, product recommendations etc. more accurate and relevant to each user. Again, the consumer benefit of such development, in terms of better search results is clear, but combined with the other platform network effects of building and retaining a wide user base, the potential to create barriers to entry is clear.

While this is normal commercial behaviour, the implied market power has two important consequences which may arise. First, it is hard for a competitor platform to enter, since they will face costs to gather the data, while the market leader receives revenue to access the same information. Second, the platform may have greater scope to set prices and commercial terms for the companies that it works with. As part of these commercial terms the platforms, since they are acting as intermediaries, can control what data flows from the buyer to the ultimate service provider. As highlighted in an earlier section, this data has considerable commercial value to the seller, who will want to know as much as possible about the buyer, for cross-sell and upsell purposes, and in these 'winner-takes-all' type markets, the barriers to entry become ever higher.

A recent European Commission fact-finding investigation into B2B marketplaces[38] found that the strong platform effects in digital business could lead to platform operators *"placing themselves in a gatekeeper position which may easily translate into (dominant) positions with strong market power"*. The report highlights that due to the data-driven nature of their business, platform operators can gain an information advantage which they can use to the disadvantage of users and found clear examples of this advantage being used by providers to set what were considered unfair terms, including restricting the data that merchants using the marketplace could access regarding the customers they were selling to.

Another high-profile case which is currently playing out involves the indirect network effects that a large platform provider can leverage by using its reach to collect additional data on its users. Germany's antitrust watchdog have issued a preliminary legal assessment with accuses Facebook[39] of collecting and processing users' data in a way which they would be unaware of by tracking Facebook users when they visit other third-party sites which contain the Facebook API, even if the user does not click on the icon. Because so many sites use the Facebook API as a means of engaging with users, Facebook can collect detailed data on the web browsing habits of its own customers, even when they are not using the Facebook service. This enhances the detail and extent of the data Facebook holds about its users. This can be processed and so it strengthens Facebook's offer to advertisers who want to target groups ever more specifically since Facebook has detailed insights about which users visit which websites.

It is important to note that it is the nature of the data value chain and data as an asset that enables these platforms to emerge. The direct network effects of such platforms deliver a clear benefit, including by potentially helping to overcome the problems and costs that would arise if one were to trade data multiple times within the value chain. In most cases for a single service, platforms are the most practical and commercially efficient way to organise activities across the value chain and the platforms that arise in many data-driven services attest to this. However, the way such platforms use the data that they collect as part of their business as a tool to control the platform and entrench their position shows that the indirect network effects are also strong and need to be considered. Having data on the transactions, behaviours and patterns of business across an entire segment may give a platform operator a significant and potentially market distorting position.

An additional effect, perhaps unique to data, is the feedback loop which can exist between data collection and analysis. As a company builds services and analytical algorithms, the insights from these can be used to reduce the cost of data collection. This in turn enables them to collect more data and further improve the algorithm. When this occurs, the cost of data collection is higher for a new platform or service provider than it is for an existing one. This effect is discussed in more detail in a CERRE report[40] which points out that the impact of this effect is linked to the cost of collecting data in the first place. If this is low to start with the feedback loop, if it occurs, will make little difference but there are instances where it could be significant.

Conglomerate Effects

Data collected as part of the provision of one service can then also be used by a company for the provision of another, unrelated service. It has always been normal commercial behaviour for a trader to look to find adjacent commercial opportunities to sell more to their existing customers, broadening their portfolio with add-on, complementary goods and services. Alternatively, companies may take skills and capabilities, such as technical expertise or a distribution network, and use this to expand into new commercial areas. Data-driven businesses are no different. Online retailers, given what they know about their customers, will very often expand their range extensively. When Google purchased the smart home devices company Nest, it was at least partly driven by obtaining access to an installed network of home sensors and the data they generate. The acquisition was rather distant from Google's core search or online services business, and Nest retained its brand and existing customer data privacy policies. However, one commentator still described it as enabling Google to learn "What we do when we're not on the web"[41]. Even anonymised home sensor data gives Google access to a new source of data and crucially an increased variety of data to enrich its existing datasets.

38. European Commission, Business-to-Business relations in the online platform environment, Final Report 22nd May 2017
39. http://www.bundeskartellamt.de, Preliminary assessment in Facebook proceeding, 19th December 2017
40. CERRE Report "Big Data and Competition Policy" 2017

The tendency for certain companies to operate in multiple markets is often cited in competition policy as the conglomerate effect. Just as in traditional industries, with large global players such as GE, Samsung and Mitsubishi, a conglomerate can be beneficial in terms of enabling companies to share common costs, diversify businesses and ease access to capital and talent. The same principle applies in data-driven businesses, but in data-driven businesses the benefits go beyond the sharing of risk and common overhead costs. Because data can be shared freely within an organisation, such companies are able to extract much greater insight from the larger data pool they have access to. In some cases this may be due to pooling data about the common user group and so developing deeper and more granular insights about those users and their behaviours than any individual player competing in only one segment would be able to achieve.

Concerns have been raised about how conglomerate companies might use their strength in one area to gain a potentially anti-competitive advantage when seeking to enter, or even just decide whether to enter, adjacent or unrelated service areas. Microeconomic literature on the effects of leveraging on competition and the resulting foreclosure effects, depending on the characteristics of the markets and goods (complementary components or standards, size of network effects, competition on endogenous innovation, etc.) has shed light on those concerns and on the way to deal with them. In the data value chain one problem is when one company acts as both a direct-to-consumer retailer and also a provider of a retail marketplace. Because it has access to otherwise confidential data on all transactions in the marketplace, including volumes and price points, it is in a very strong position to decide when to add certain products to its own retail portfolio and be certain of the margins it will make when doing so. Facebook has used data from a mobile analytics app (which many consumers download to monitor their own data usage) as an early warning tool to identify new apps that may quickly take-off to become potential competitors enabling it to counter and/or try to acquire the new player. All such companies must of course abide by anti-trust law prohibiting abusive behaviour but the potential for conflicts of interest remains high.

Amazon, as well as being a very successful online retailer, has become a powerful logistics company with a home delivery network which it can fill with its own deliveries but also those of third party sellers. It has also vertically integrated from e-commerce to making its own home electronics, launching a streaming video service and recently via its acquisition of Whole Foods has moved into bricks and mortar retailing. Its ability to bundle the sale of these services and track user

41. Wired Magazine, 14th January 2014, What google really gets out of buying nest for $3.2 billion,

behaviours and preferences across the multiple services is significant. The policy concerns raised in public about Amazon so far have mainly been around its advantages in tax and the pressure on labour and other input costs, but over time there will likely be heightened attention to competition policy aspects and the extent of its influence in retail and B2B markets.
A massive e-commerce conglomerate which gained a greater profile outside China after it filed its IPO in September 2014 is the Alibaba Group. The company runs several marketplaces, cloud storage businesses, a search engine and financial services through an independent subsidiary, Ant Financial Services Group. Through this they also operate an online bank as well as China's most popular online payment service. The breadth and volume of services, and the ability to pool deep and hard-to-replicate data about customers and suppliers, creates a strong market position. Conglomerate effects can thus be national (if the home market is very large) and not just global.

There are concerns that because current competition frameworks focus on consumer benefit and primarily measure this using price, competition authorities may be missing the real extent of anticompetitive effects of data-driven business models. In a paper published in the Yale Law Journal[42], Lina Khan argues that the current antitrust approach "is unequipped to capture the architecture of market power in the modern economy. [.....] Specifically, current doctrine underappreciates the risk of predatory pricing and how integration across distinct business lines may prove anticompetitive". The article also highlights the particular issue of platform providers who compete directly with users of their platform, *"This dual role also enables a platform to exploit information collected on companies using its services to undermine them as competitors"*.

Privacy and Trust

Trust and confidence in data-driven services and their providers is a key concern for users, and a key objective for policymakers. In protecting privacy, the aim is to build trust and confidence that private data are being adequately protected according to applicable privacy regulations and requirements. This requires all parties involved to adopt a coherent approach that is technology neutral and consistent across all services, sectors and geographies. In the context of the mobile industry, the GSMA has developed a set of Mobile Privacy Principles, which describe the way in which mobile consumers' privacy should be respected and protected when they use mobile applications and services that access, use or collect their personal data. The key overarching objective of these principles is to foster business practices and standard that deliver meaningful transparency, notice, choice and control for mobile users with regards to their personal information and the safeguarding of their privacy. These principles seek to strike a balance between protecting an individual's privacy and ensuring they are treated fairly while enabling organisations to achieve commercial, public policy and societal goals.

One of these principles - accountability and enforcement - emphasises that organisations adhering to the principles must be accountable for ensuring that the principles are met. The principles of accountability and effective enforcement also underpin the APEC Privacy Framework[43] and the APEC Cross Border Privacy Rules[44] system. In Europe, a principle-based, risk-based approach is at the centre of the General Data Protection Regulation (GPDR), which clarifies the basis on which data can be collected and conditions of when and how it can be used. However further obligations may be imposed to telecom operators by the future EU ePrivacy regulation. In some countries, individuals can request to see the data that companies hold on them and have various rights to amend or request its removal in specified circumstances. In practice, using these rights today involves a complex and time-consuming process for individuals. The principle of accountability and enforcement, in this context, becomes a key to safeguarding consumers' privacy. Finally, the expected rapid growth in the Internet of Things is bringing new security and trust issues to the fore. It seems likely that such issues will have high prominence in public debate for the foreseeable future. The GSMA has adopted several policy positions to inform this discussion and strike the right balance between consumer interests and commercial innovation, as mentioned earlier.

42. Khan, Lina M, Amazon's Antitrust Paradox, Yale Law Journal 2017
43. file:///C:/Users/sabate/Downloads/05_ecsg_privacyframewk.pdf
44. http://www.cbprs.org/

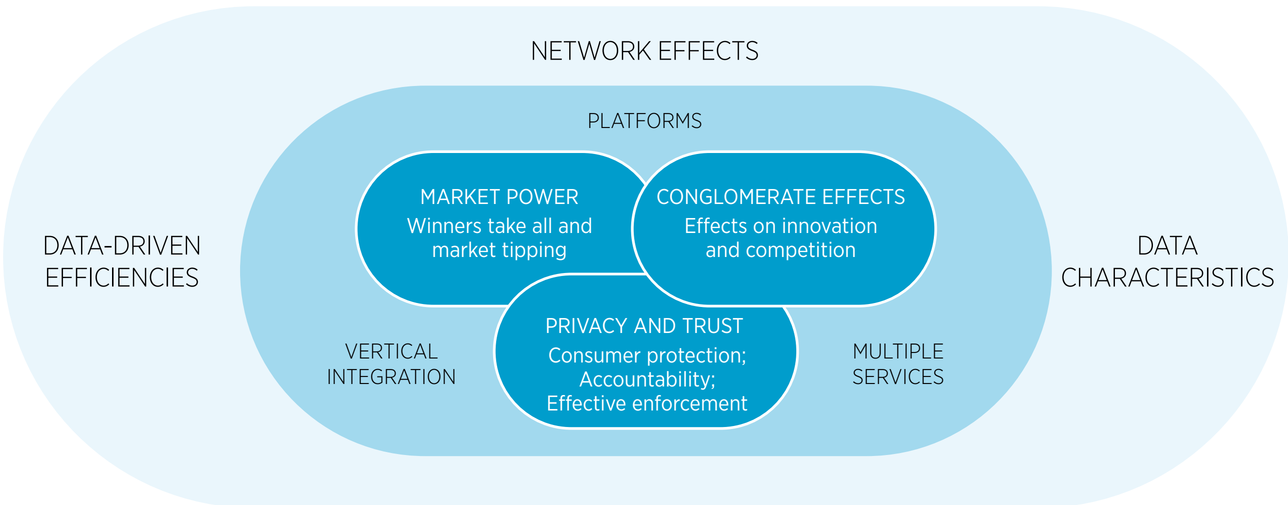
Policy Implications

As has been highlighted throughout this report, there are instances where the data value chain could work better, with opportunities for competition to be increased.

It is generally accepted that new or evolving markets should enjoy a degree of policy leeway to ensure that innovation can flourish. Nonetheless there are policy issues which either pose a barrier to optimal competitive outcomes in today's value chain or which merit close attention for the future. The analysis of the value chain and the policy environment point to three areas to monitor.

Consistency in Rules

From a business perspective, the most obvious concern is fragmented and inconsistent regulation – between geographies, sectors, and within sectors – which creates operational challenges for global firms and uneven playing fields that impede market competition. In the data value chain, different players are subject to different regulations about the same issue. For example, while telecom operators and social media companies may collect similar data on user location, either through mobile network location-based services or GPS information via the phone, the latter are able to analyse and commercialise such location data, whilst telecom operators are typically restricted by sector-specific regulations from doing so as freely. In most jurisdictions they face specific additional requirements that include a higher standard of consent before they can use metadata about users. Similarly, many telecom operators have specific license obligations (which vary for each country they operate in) to store and process data within the country and requirements to co-operate and assist local security authorities which do not always apply to internet businesses These requirements add additional cost in terms of reduced economies of scale in markets which are often global in scale and restrict freedom to act in a normal commercial way. Even more importantly, they restrict competition in data markets when policymakers should be looking to promote the widest possible competition. To create a fair, competitive and efficient market, policymakers need to address these disparities by removing the restrictions from the more regulated players, so that all players have an equal chance to compete, and consumers have clarity on what to expect. The obvious remedy would be to remove sector-specific regulations in areas where the same or substantially similar type of data is being processed.



Market Power and Network Effects

Many data-driven businesses and platforms have strong direct and indirect network effects which are drivers of efficiency and beneficial for users. Vertical integration and closed business models have a place since they can facilitate the launch of new services where closer integration of device and platform are needed and standards have not yet emerged. However, in data-driven businesses the direct network effects can lead to especially strong market positions and indirect network effects can further strengthen positions and become in effect very high barriers to entry for any potential competitor. In fact, for some of the largest internet platform-based businesses, they already have so many users and hold so much data about these users, that even with massive investment it is difficult to imagine a new entrant being able to compete effectively in creating the necessary virtuous circle to attract users and advertisers.

It would seem risky to adopt an ex-ante, sector-specific regulatory approach to the potential data market power issue. This ex-ante approach has caused problems in mature sectors like telecoms, and in a new, evolving sector could potentially do more harm to innovation than good. This does not mean that a purely laissez-faire approach is appropriate either. While perhaps justified in the early days of the Internet, the markets where data is highly relevant represent too high a concentration of economic power to impose an "exclusion zone" for policy. The preference should be timely, well-targeted ex-post interventions to stop demonstrated and durable abusive behaviour, having considered dominance beyond traditionally defined markets as well as the impact on innovation. While ex-post action is typically the best approach, given the dynamism of the market, this should proceed quickly.

One challenge competition authorities face is being able to identify to what extent the undoubted benefit of network effects to consumers can be offset by the long-term costs of monopoly. 'Traditional' monopolistic behaviours such as excessive pricing, may manifest themselves in data-driven markets in a different form, such as too much data being requested, so that authorities need to be just as alert to these potential

behaviours and the impact on consumers and businesses. This implies that the routine "toolkit" of competition policy is difficult to deploy, since there is not yet a transparent market with clear pricing/revenue indicators for each value chain activity (in fact some services may be zero-priced to customers as the data is being marketed to a different customer). This lack of transparency makes it harder to identify, investigate and remedy abuses of market power should they occur. Given these complexities, policymakers need access to up-to-date research and tools so that they can act early enough. Further research is needed on the economic benefits of platforms and on how companies may use or abuse market power as these evolving business models are less well understood and so hard to challenge. Antitrust policy also needs to focus more attention on the interactions across apparently distinct markets given the steady expansion by internet platforms and tech companies into other activities. They need to take a forward-looking view of converging markets and the conglomerate effects, recognising the additional value and competitive advantage companies gain from playing in and leveraging data, insights, sales power across multiple segments, rather than a retrospective view of narrowly defined sub-markets.

Access to Data

Issues of access to data and the rights to use it are becoming increasingly important and are distinct from the issues of privacy and security. These are being dealt with separately by policymakers in many jurisdictions through both new legislation and the creation or strengthening of data ombudsmen and enforcement authorities. As data analysis techniques become ever more sophisticated, greater focus will be demanded on how the data is analysed, who has rights on the output of such analysis, and to what purposes that output can be put.

Here too there would be a risk that too much ex ante regulation could block innovation or simply create temporary distortions that technology trends or regional differences render redundant. There is a much stronger case for self-regulation across the value chain, with codes of conduct and commitment to transparency and open debate that effectively address customer concerns and provide guiderails for all players. Ideally there should be as much interoperability as possible across regions and sectors on this question, but it will be necessary to respect established cultural differences that prevail across the globe.

There is some debate about whether value chain players should be obliged to open their data to others, to advance innovation and competition. Where the data in question has been generated by public bodies or funded by taxpayers, openness may be a reasonable intervention to consider. In the commercial space, the potential downside of reducing incentives to invest in data-driven businesses (whether generating, collecting or processing data) would caution against significant interventions unless specifically required, e.g. ex-post interventions or merger remedies. Where value chain players have invested in the generation, collection and secure storage of data valuable to their business operations, they have a legitimate right to protect their exclusive right to use that data while respecting the applicable rights of customers and business partners and in line with competition policy.

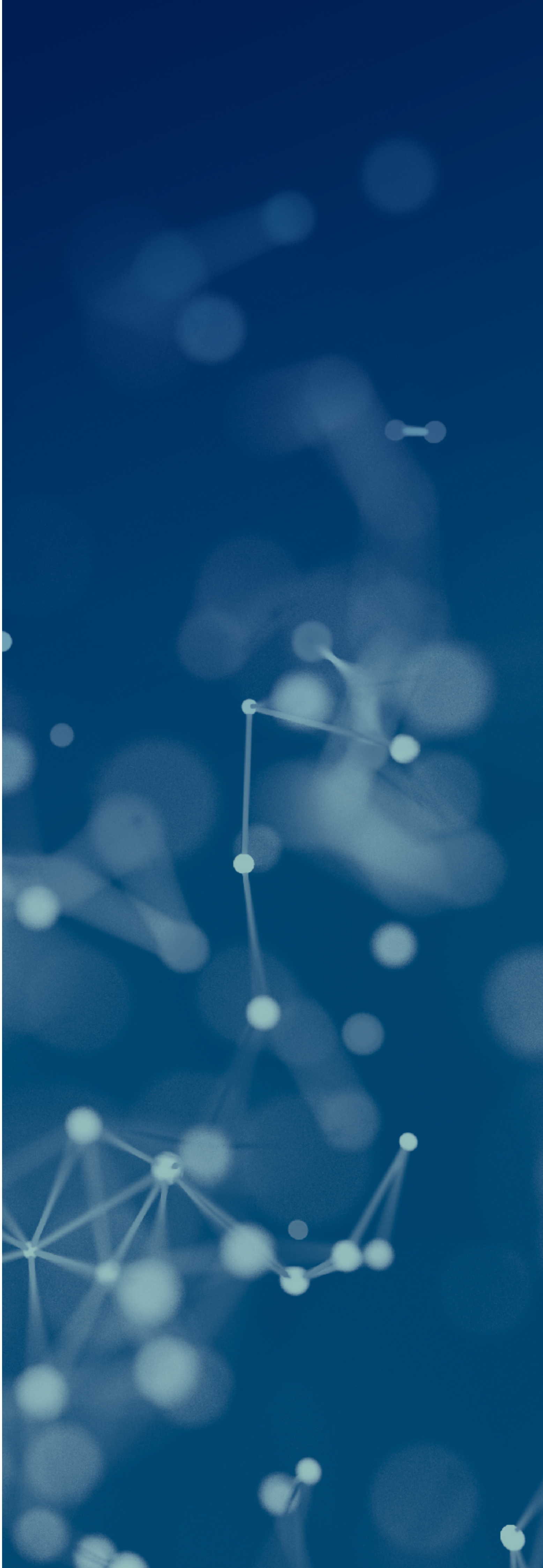
Conclusions

The data value chain is important in almost all industry segments and the value that is generated through the ability of organisations to collect, store and process data is self-evident. The rapid growth of data-driven business models and the anticipated continued growth in automation, digital technologies and artificial intelligence, running on ever more powerful processing and storage infrastructures is a clear sign that this trend will continue. It is also clear that while this growth has been strong to date, it has not been without issues that concern policymakers keen to promote continued competition, and an efficient and fair value chain which serves consumer interests as best as possible.

The nature of data and its characteristics as an asset make it relatively hard to trade, leading to integrated business models where companies seek to collect source data in order to be able to process it through the various steps of the value chain. While the consumer benefit of these data-driven businesses in terms of low prices and free services has been clear, there are growing concerns about the effects on competition. The current market competition tests being based on consumer outcomes in terms of price only may no longer be suitable for data-driven services, as other factors, such as privacy, play a vital role in the value exchange between customers and service providers.

It is also clear that competition in these services needs to be governed by consistent rules (including consumer protection) and that any sector-specific rules that apply to a subset of competitors have the potential to distort competition and should be removed if they give rise to barriers to open competition.

In summary, current competition and regulatory approaches may not be suitable for the unique characteristics of data-driven business. More work is required to adapt these approaches to protect the very wide set of policy objectives. At the same time any policy regime must above all ensure that the data value chain continues to thrive as major driver of economic growth and social progress.



GSMA HEAD OFFICE
Floor 2
The Walbrook Building
25 Walbrook
London EC4N 8AF
United Kingdom
Tel: +44 (0)20 7356 0600
Fax: +44 (0)20 7356 0601