QUINLAN & ASSOCIATES

THE REAL LINGUA FRANCA (PART 1/2)

UNLOCKING THE POWER OF DATA

THE AUTHORS

BENJAMIN QUINLAN
CEO & MANAGING PARTNER
BCom (Hons 1) / LLB (Hons), *Macquarie University*

JUSTIN CHUNG
ASSOCIATE
BBA (Hons 1), *Hong Kong University of Science and Technology*

CHARLES YAU
SENIOR CONSULTANT
BBA (Hons), *Hong Kong University of Science and Technology*

CONTENTS

EXECUTIVE SUMMARY 4

SECTION 1: DATA: BEHOLD, MORE GOLD 6

SECTION 2: DATA 101 13

SECTION 3: DATA VALUE CHAIN 23

SECTION 4: CONCLUSION 47

SECTION 5: HOW CAN WE HELP? 49

EXECUTIVE SUMMARY

Data has been trumpeted as the new gold for the 21st century more times than one could possibly count. And it does bear similarities to its physical cousin. Both are malleable, with the ability to be forged and manipulated into something of tremendous value in the hands of a skilled craftsman or data scientist. Both are also highly reflective; gold is used in the visors of astronaut's space helmets to reduce glare and heat from the sun, just as data is used to reflect errors in our thinking to reduce possible human biases in our work. Even in its most raw form, gold is prized for its beauty, while data is prized for its simple face-value insights.

However, there is one key difference between gold and data: abundance. Gold is one of the ten rarest elements on the Earth's crust, with mining consolidated in only a handful of places across the globe. Data, on the other hand, has become so abundant that you wouldn't even notice it being mined right under your nose.

In recent years, we have witnessed the explosion of data sources; from the smartphones in your pocket to the smallest IoT sensors. With it, the amount of data generated has jumped from 13 Zettabytes in 2013 to 53 ZB in 2019. This abundance of data has created billion-dollar companies in the data economy and created new opportunities for businesses worldwide. However, these eye-popping valuations have also created significant fears of missing out ("FOMO") for executives looking to leverage their company's data to gain a competitive edge in the market.

With data usage terms like "artificial intelligence", "data analytics", and "automation" being thrown around on an almost daily basis, it is easy to get caught up in what data *could* do as opposed to what it *can* do. It is important that executives not get distracted by lofty outcomes of using data but instead understand what data is available to them, what it conveys, and how it can support current business strategies. In keeping with the gold analogy, an expert goldsmith can only craft the best jewellery if they fundamentally understand the gold nugget in their hand; its properties, limitations, possibilities, and how to manipulate or augment it to fulfil a specific purpose.

Like smithing, there is a process to harness the full potential of data, known as the "data value chain." It specifies the complete lifecycle of a nugget of data: from data collection to storage, usage and, finally, disposal. All four stages are governed by a robust data centric culture, cybersecurity measures, and regulations, which are changing rapidly across all jurisdictions.

Despite all the hype, we believe there is significant wastage in the market around investing in data-centric projects. In particular, we find many companies are spending significant amounts of capital in the "usage" stage of the data value chain without a firm understanding of what it genuinely means for their business or how it falls within the company's broader data strategy. In reality, these half-baked data projects create considerable losses in investment or, even worse, fail to meet desired business outcomes. We estimate that more than 95% of companies globally, excluding micro unlisted firms and NGOs, do not have a complete grasp of the data value chain within their organisation, leading to silent but substantial losses.

In our two-part report series, we look to: (1) unpack the data value chain; and (2) showcase what a holistic data strategy project looks like, incorporating both business and technology elements to ensure executives structure data projects around their core business strategy, such that every dollar spent on transformation or optimisation is put to good use. After all, data is – and always will be – the real lingua franca of business.

REPORT STRUCTURE

Before deep-diving into the ins and outs of a data strategy project, it is important to understand: (1) basics of data and the supporting infrastructure and (2) data value chain - from collection to disposal.

The report is structured in three different sections, covering topics at different levels of data understanding. Based on your understanding of the data space, please feel free to skip to the next paper to immediately dive into data strategy projects proper.

95% OF COMPANIES GLOBALLY… DO NOT HAVE A COMPLETE GRASP OF THE DATA VALUE CHAIN WITHIN THEIR ORGANISATION, LEADING TO SILENT BUT SUBSTANTIAL LOSSES

SECTION 1
DATA: BEHOLD, MORE GOLD

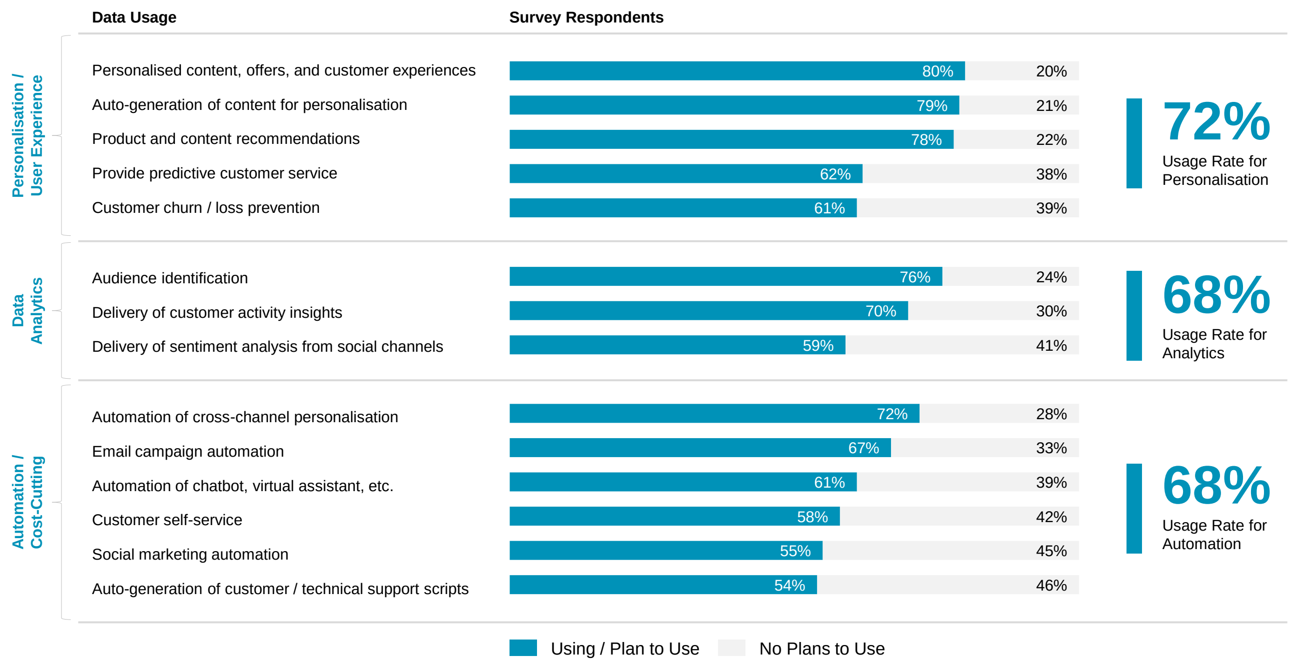
INTRODUCTION

For many of us, a typical day involves being woken up by a smartphone alarm, then grabbing our phones to read through personalised news updates, check work emails, and browse through endless social media feeds…and perhaps even play the occasional game of Candy Crush. Whether we like it or not, technology has become inextricably linked with our daily lives, both in and out of the office.

For much of the 20th century, companies offered mass-market, one-size-fits-all products designed to maximise consumer coverage. Newspapers used to include a comprehensive suite of topics, ranging from politics to lifestyle. However, rapid technological developments have resulted in hyper-personalisation, enabling companies to customise their products based on each consumer's specific interests and needs. As such, readers no longer need to dig through pages of news to find relevant content – articles are now curated based on reader preferences and then pushed for consumption.

The truth is every piece of activity data or digital footprint created during application usage is tracked and collected by software providers. This data is then analysed to better understand user behaviours to push more relevant content and recommend targeted products. As a result, the more an app is used, the more addictive it often becomes due to personal relevancy. Indeed, product personalisation / customisation is one of the more prevalent use cases of data across industries, with 72% of companies leveraging data to maximise user engagement and minimise customer attrition (see Figure 1).

FIGURE 1: HOW DATA IS USED ACROSS INDUSTRIES



Source: Adobe, Quinlan & Associates analysis

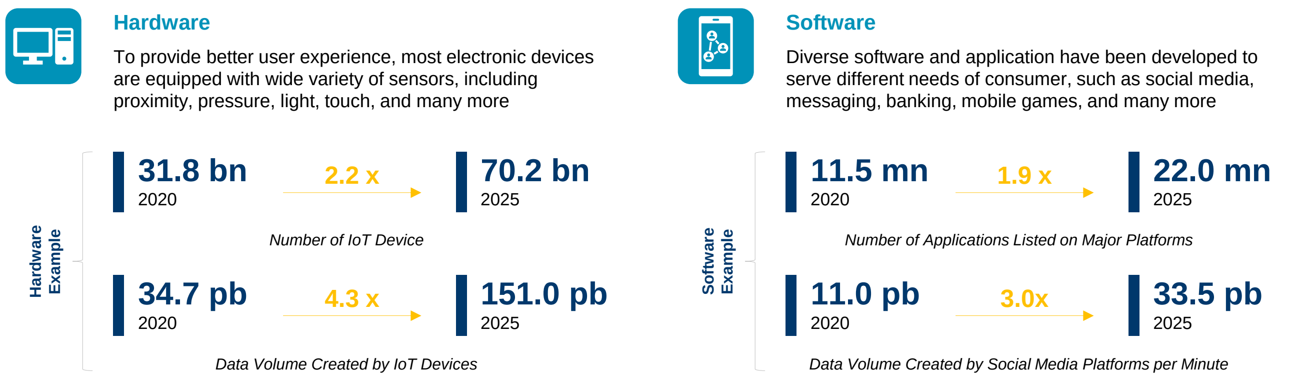
In addition to personalisation, companies also use data for (1) analytics to better understand their target audience, and (2) automation, for cost-reduction purposes.

With sufficient analytics, data creates a self-reinforcing cycle between companies and consumers, with consumers generating data during consumption, which is harvested by companies to drive engagement levels and further consumption, which in turn creates more data.

1. DATA SOURCES

With ongoing technological advancements and increasing adoption, especially in the wake of COVID-19, the rapid pace of hardware and software development and deployment is forecast to grow in its use and size over the coming years (see Figure 2).

FIGURE 2: DATA GENERATION



Source: HIS Markit, IDC, Seagate, CISCO, Quinlan & Associates estimates

1.1. HARDWARE

There were 31.8 billion Internet-of-Things ("IoT") devices as of 2020, with the number expected to more than double by 2025, reaching 70.2 billion. To put this number into perspective, if 70.2 billion iPhones are stacked up on top of each other, the height of the tower would reach ~523,000 km, which is 1.35x the distance to the moon.

These IoT devices, such as smartphones and TVs, are equipped with a wide variety of sensors to track and record a diverse range of data, including distance / proximity, pressure, and brightness. In 2020, an estimated ~34.7 petabytes ("PB") of data was created by these devices per minute, with the volume expected to grow 4.3x to 151.0 PB per minute by 2025. To put this into a perspective, 1 petabyte is 1,000,000 gigabytes. Assuming aggressively that an average smartphone user consumes 10 gigabytes of data a month, it will take the user 8,333 years to reach 1 petabyte. To reach ~26 petabytes, which is the amount of data created per minute, that will take approximately 217,000 years.

1.2. SOFTWARE

With mass adoption of broadband and mobile internet driving ongoing advancements in software development, the number of applications listed on major platforms (such as Apple App Store) has exploded in recent years. These applications were developed to address specific user needs, with the top five categories being games (21.5%), business (10.1%), education (8.7%), lifestyle (8.6%), and utilities (6.3%).[1]

There were over 11.5 million applications available on major platforms as of 2020, which is expected to grow by 1.9x to reach 22.0 million by 2025. These applications gather activity data and content created by users, which can be monetised using various methods, including sales acceleration via the recommendation of relevant products / services and third-party advertisements.

The amount of data generated and captured by applications is enormous – social media platforms alone generated 11 PB of data per minute in 2020. To put this into perspective, if one were to use 10 GB of data per day, it would take over 3,000 years to reach 11 PB of data.

It is also worth noting that paid applications make up only ~4% of all listed applications. With ~96% of applications being free or freemium services, this clearly demonstrates the power of data; the value extracted from data can be monetised and duly compensated for the cost of providing and running these applications.

SOCIAL MEDIA PLATFORMS ALONE GENERATED 11 PB OF DATA PER MINUTE IN 2020… IF ONE WERE TO USE 10 GB OF DATA PER DAY, IT WOULD TAKE OVER 3,000 YEARS TO REACH 11 PB OF DATA

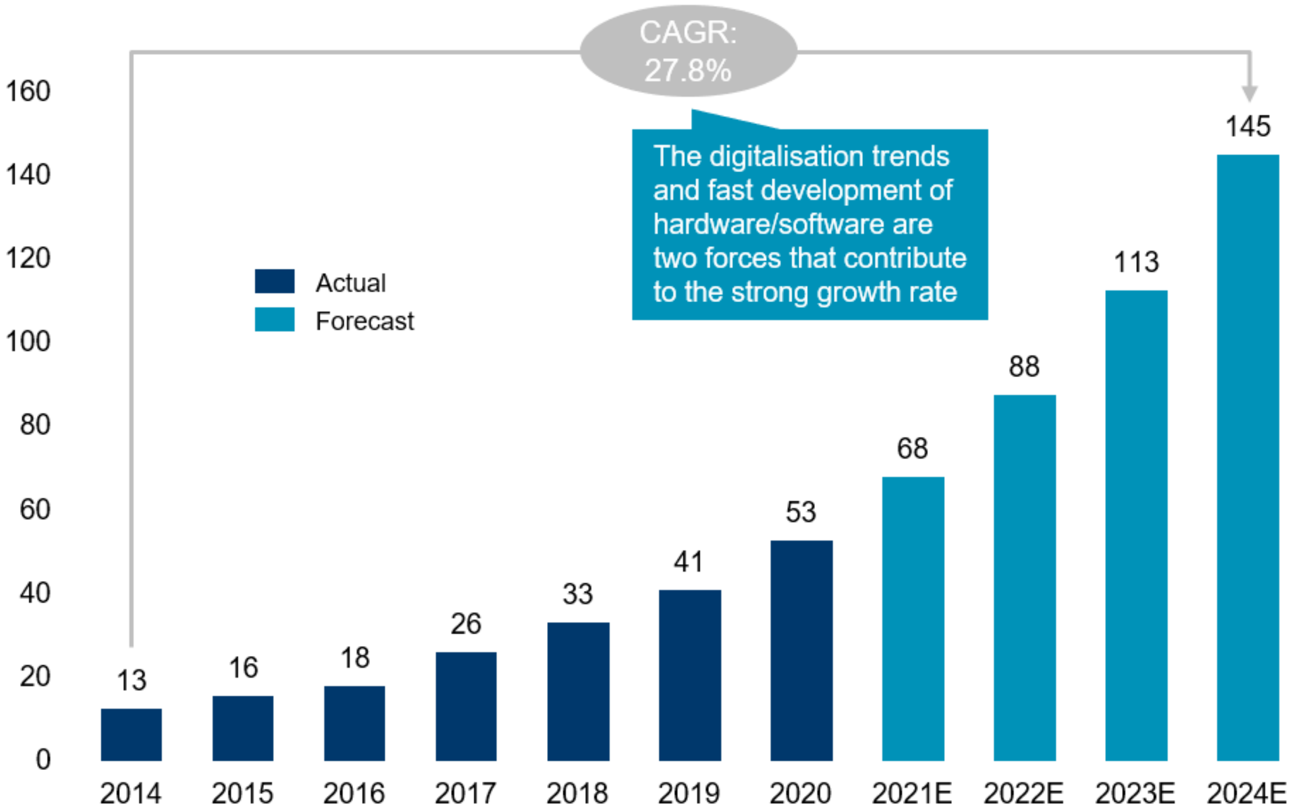
1. Apple App Store, 2021

2. DATA VOLUME

Moore's Law, a widely accepted concept until recently, theorises that the number of transistors on a circuit should double every two years. By its strictest definition, Moore's Law is no longer happening, though there will eventually be workarounds or advances in technology which will continue to drive the exponential rise in data processing power and, directly, how much data is generated and collected. Whether it is through new materials like graphene and nanomagnets or processes like quantum computing, data will continue to be ubiquitous. In other words, information processing power has been – and is expected to continue – improving rapidly.

Assuming current trends persist, and with increasing processing power, we expect the growth trends in data volumes to persist, growing by ~30% p.a. to reach 145 ZB by 2024 (see Figure 3).

FIGURE 3: DATA VOLUME



Source: IDC, Quinlan & Associates estimates

3. DATA MONETISATION

The fact is, the world is now driven by data, and the importance of it will only increase over time. For companies of all industries, understanding how to effectively utilise data is no longer a nice-to-have, but fundamental to future survival.

A growing number of examples of this have surfaced in recent years, especially in the internet space – the most notable of which include social media giants and internet service providers who sell customer metadata data to 3rd parties. However, more traditional industries are beginning to incorporate data into their books as a core stream of revenue (see Figure 4).

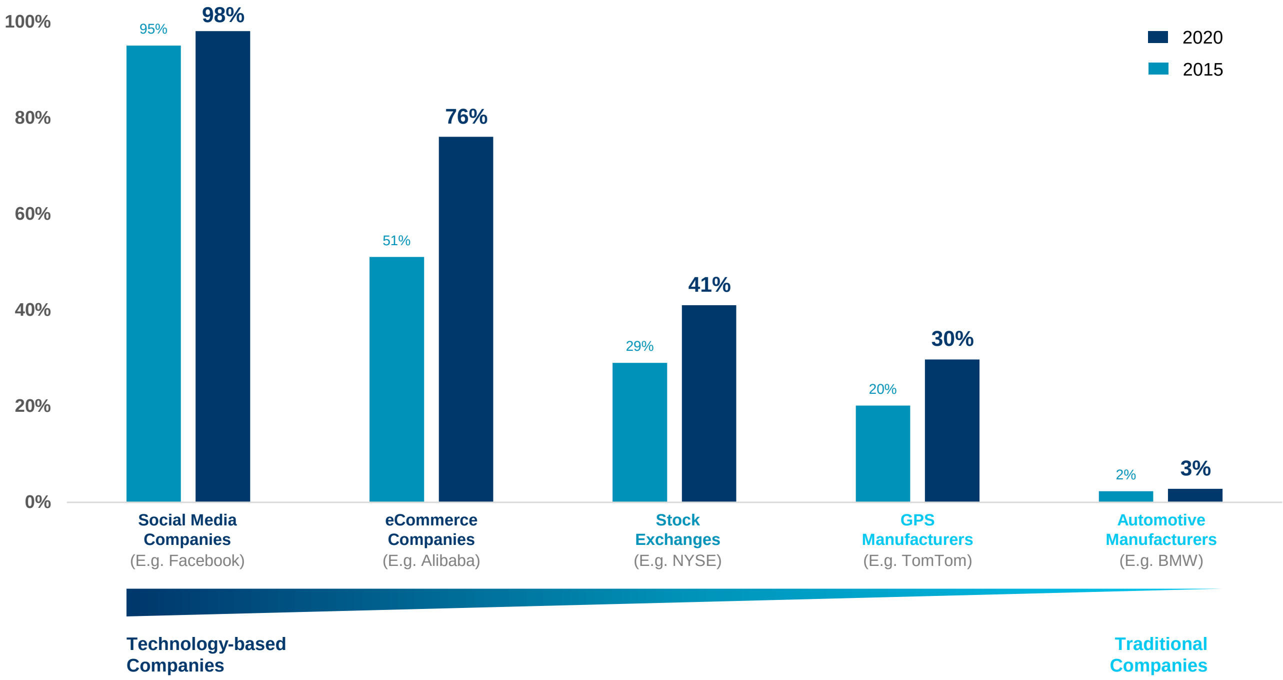
Take stock exchanges, who have traditionally relied on their core trading and listing fees to drive their revenues. With heavier global competition and strong demand for trading data, leading global exchanges, including NASDAQ, London Stock Exchange, and NYSE have focused on developing their in-house data capabilities, generating considerable revenues from the sale of such information. Today, data services generate up to 42% of total revenue for these exchanges.

Traditional GPS manufacturers, crushed by the advent of tiny GPS modules in smartphones, have also pivoted their business models, selling location data in automobiles to 3rd parties. Noticing this, other hardware-focused industries such as automotive manufacturers are taking their first foray into the data market, capitalising on their increasingly connected vehicles, which are generating up to 25 gigabytes of data per hour.[2]

UNDERSTANDING HOW TO USE DATA IS NO LONGER A NICE-TO-HAVE, BUT A FUNDAMENTAL TO FUTURE SURVIVAL

[2]2. Washington Post, “What does your car know about you? We hacked a Chevy to find out.”, available at: https://www.washingtonpost.com/technology/2019/12/17/what-does-your-car-know-about-you-we-hacked-chevy-find-out/

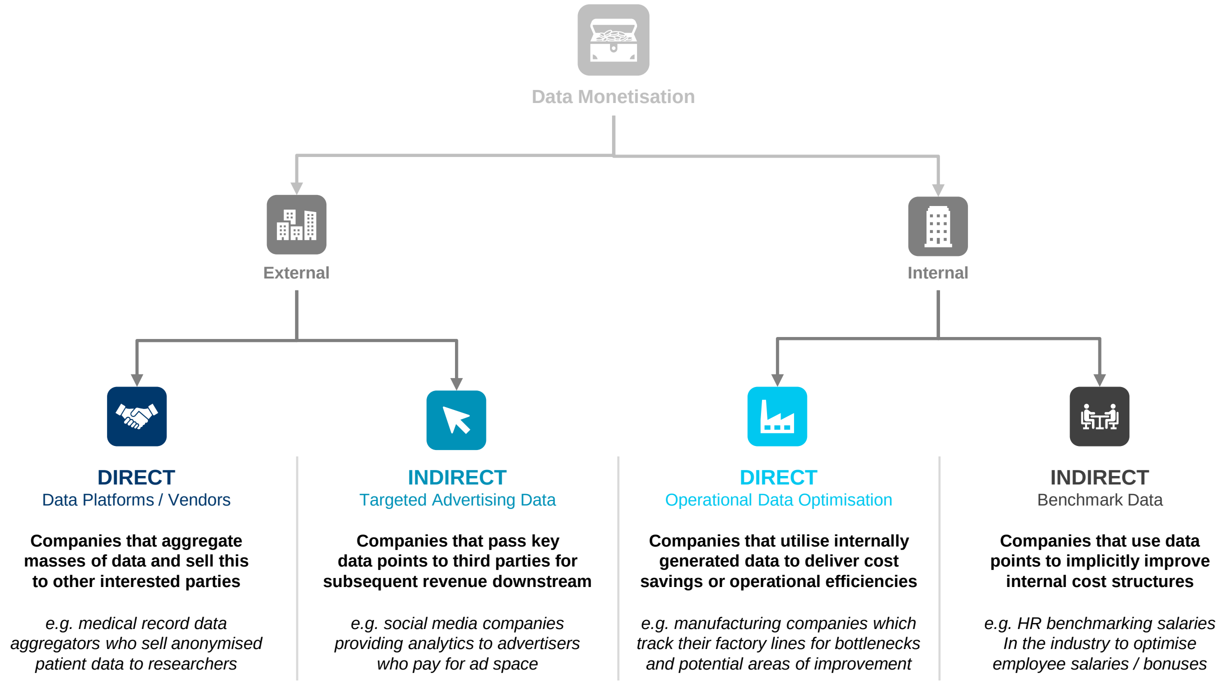
FIGURE 4: DATA SERVICES AS A PERCENTAGE OF TOTAL REVENUE



Source: various company and industry annual reports, Quinlan & Associates research and analysis

It is important to remember that there are a variety of ways for companies to monetise data, some less obvious than others (see Figure 5). Firms can look externally: generating revenues from customers directly via the sale of data on hand or indirectly via downstream revenue sources. Companies can also seek to drive revenue growth internally: leveraging data to optimise internal processes (e.g. production or employee compensation). Whichever path companies choose, insights derived from data are increasingly being used to deliver tangible financial outcomes.

FIGURE 5: DATA MONETISATION MODELS



Source: Quinlan & Associates research and analysis

As more companies are recognising the ability to convert their data into actual gold (or green, as it were), it is of vital importance that executives understand data fundamentals and how it can be used to shape their businesses.

In this first part of our two-part series, we hope to equip any executive with the ability to craft their organisation's data journey with utmost confidence.

SECTION 2
DATA 101

Everything we see, hear, or feel contains information. However, the information must be captured and processed to be understood meaningfully. Similarly, data is simply a piece of information that is stored in a “raw” form. The raw material needs to be processed before it can be utilised in a practical manner.

1. DATA FORMS

Raw data typically comes in one of two forms: quantitative or qualitative. While both forms are useful and valid depending on the intended use, they differ in cost, ease of use (and collection), and value.

More recently, a “third” form of data has emerged as a result of technological advances in artificial intelligence and machine learning: quantified-qualitative data (see Figure 6). Quantified-qualitative data attempts to extract numerical values from qualitative information, capturing the insightful value of qualitative data but proxying it in quantitative form so that it can be analysed using mathematical formulae and functions.

FIGURE 6: FORMS OF DATA

	123 QUANTITATIVE	ABC QUALITATIVE	A=1 QUANTIFIED-QUALITATIVE
Description	• Purely numerical data, and likely the most familiar form for many	• Any raw source of information that is not in a numerical form	• Qualitative data that has been transformed into numerical form
Examples	• Market data • Government census statistics	• Images / Videos • Text and Voice Recordings	• Qualitative data processed and tagged by data scientists / software
Cost	✓	-	✗
Ease-of-Use	✓	✗	✓
Value	-	-	✓

✓ Favourable   - Dependent   ✗ Unfavourable

Source: Quinlan & Associates research and analysis

1.1. QUANTITATIVE DATA

Quantitative data is presented in numeric format and is likely the most familiar form of data for many people. Examples including financial market data and government census statistics.

Quantitative data is relatively easy and cheap to collect and prepare for analysis. Many companies generate the data they require, such as consumer statistics and sales data, through their daily operations, making it mostly free to acquire. Moreover, the evaluation of small quantitative datasets is extremely common, with one of the most widely used evaluation tools being Microsoft Excel.

However, because of the accessibility and simplicity of quantitative data, the results obtained – and hence the value created – tend to be relatively generic and commercialised. As such, the analysis of quantitative data by companies is: (1) more of an operational necessity rather than a competitive advantage; and (2) may overlook vast amounts of information that could be captured and utilised in other forms.

1.2. QUALITATIVE DATA

Qualitative data is any raw source of information that is not in numerical form, such as free-text survey responses, photos, voice recordings, and speech transcripts.

Qualitative data is typically more expensive to collect and store than quantitative data. Depending on its format, additional equipment or personnel is often required to generate such data. For example, cameras are required for images and videos, while individuals are needed to transform speeches from an audio form into textual form. Furthermore, because of its nature, qualitative data tends to be larger in size than quantitative information, requiring more digital space – and hence cost – for storage.

Objective analysis evaluates a situation based on a pre-defined scale of measurement. However, as qualitative data is typically unstructured (and hence lacks an inherently and naturally quantifiable value), it is harder to evaluate it in a consistent manner. Therefore, analyses of such data typically result in suboptimal, flawed, or biased conclusions, making it harder to use.

Many companies incorporate qualitative analyses into their data evaluation process. Even though the implications cannot be quantified, such analyses can provide directional implications. For example, “this product is bad” provides a negative indication of the product, despite not giving any quantifiable indication of how bad the product is. While insights can be extracted from qualitative data, more needs to be done to maximise its value.

1.3. QUANTIFIED-QUALITATIVE DATA

Recognising the weaknesses inherent with qualitative data, considerable efforts are being made to design formulae or algorithms to quantify it. For example, speech transcripts can be analysed using natural language processing (“NLP”) techniques to assign a numerical score that indicates the sentiment of the speaker.

For this to happen, qualitative data needs to be collected, and then quantified, in a consistent and unbiased manner. In the past, data scientists and statisticians were hired to quantify such data by applying data analysis or statistical techniques. Nonetheless, because of the subjective nature of human beings, quantified values may not be completely accurate and can vary substantially based on techniques that are applied.

Machine learning techniques have increasingly been adopted to prepare such data. By utilising machines, which can be unbiased, and continuously improving the algorithm via feedback and reviews, algorithms can generate standardised numerical values for qualitative information. Nonetheless, this process is significantly more expensive to carry out than simply collecting qualitative or quantitative information. However, because of the numerical nature of quantified-qualitative data, it is essentially as easy to use as quantitative data, as traditional analytical techniques can be applied.

Because of the cost and sophistication required to quantify qualitative data, only organisations with sufficient resources can extract value from such information, making the insights significantly more valuable than quantitative or qualitative data alone.

2. DATA TYPES

Data can be measured against different scales. Traditionally, qualitative data is collected on a nominal or ordinal scale, while quantitative data is measured based on an interval or ratio scale (see Figure 7).

FIGURE 7: DATA MEASUREMENT

	NOMINAL	ORDINAL	INTERVAL	RATIO
Description	• Categories without any numerical meaning, essentially just labels	• Categories with an order, which could be used to be ranked	• Categories with known differences between ranks	• Categories on which basic arithmetic can be conducted
Characteristics				
Categories	✓	✓	✓	✓
Ranking	✗	✓	✓	✓
Measurable Differences	✗	✗	✓	✓
Absolute Zero	✗	✗	✗	✓
Example	• Colours • Country • Gender • Race	• Social status • Level of Education • Level of Familiarity • Level of Satisfaction	• Celsius temperature • Fahrenheit temperature • Curved Grading • IQ Test	• Height • Time • Price / Cost • Bank Account Balance
	QUALITATIVE DATA		QUANTITATIVE DATA	

✓ Applicable ✗ Inapplicable

Source: Quinlan & Associates research and analysis

2.1. NOMINAL DATA

The nominal scale is the most elementary type of scale and can simply be treated as "labelling". Data can be categorised into different classes, but there is no numerical significance to the classes.

Examples of nominal classes include colours and countries. While products can be classified by colour or by country of origin, there is no natural comparison between each class.

2.2. ORDINAL DATA

The ordinal scale is a ranked scale, in which there is a natural ordering for the classes. However, the differences between classes are not well defined.

An example is the level of satisfaction towards a product or service. Potential data points may consist of: (1) very unhappy; (2) unhappy; (3) slightly unhappy; (4) neutral; (5) slightly happy; (6) happy; and (7) very happy. It is clear that (1) very unhappy is the worst and (7) very happy is the best. However, it is nearly impossible to accurately quantify the differences between categories. For example, one cannot say (7) very happy is six units better than (1) very unhappy.

2.3. INTERVAL DATA

Interval scales are more advanced than ordinal scales, in the sense that differences between classes can be measured exactly. This provides a better sense of how categories rank against each other.

The temperature in Celsius is an example of an interval scale. Degrees are ranked against each other (e.g. 100 C is higher than 50 C), and differences between data points can be measured exactly, (i.e. there is 50 C of difference between 50 C and 100 C).

2.4. RATIO DATA

In addition to possessing all qualities of interval scales, ratio scales also have a true zero point. The nature of ratio scale enables arithmetic calculations of the data points, and hence ratio scale is the most sought-after type of scale for any statistical analysis.

Note that in the example of Celsius temperature, the scale does not have a natural zero. Even though 100 C is 50 C more than 50 C, 100 C is not twice at hot as 50 C.

An example of a ratio scale is time duration. Differences between different durations can be calculated (e.g. 100 minutes is 50 minutes more than 50 minutes), and there is a natural zero (i.e. 0 minute, or no time). Because of the existence of the natural zero, arithmetic operations (and exact comparison) can be conducted (e.g., 100 minutes is twice as long as 50 minutes).

3. STRUCTURED AND UNSTRUCTURED DATA

Data is found in two main formats: structured and unstructured. Structured data, normally quantitative, is more easily defined and searchable than unstructured data, which is normally qualitative.

3.1. STRUCTURED DATA

Structured data is easy to analyse as it exists in a predefined format. Quantitative data is almost always considered structured, and thus, exists in text and numbers such as comma-separated values (CSV) and Extensible Markup Language (XML). However, only 20% of all enterprise data is structured data.[3] As it is already processed, structured data fits in the format of relational databases or in data warehouses. This data format resides in a fixed field within a file or a record, making it more accessible through databases using Structured Query Language (SQL).

This form of data can be analysed through regression, classification, and clustering of data based on specific attributes. An example is the data stored in Microsoft Office Excel files that are organised by tables, clusters of information, with predefined connections. The data in Excel is easily accessible and filtered through leveraging the functions related to the rows and columns.

3.2. UNSTRUCTURED DATA

3. mongoDB, Unstructured Data, available at: https://www.mongodb.com/unstructured-data

Unlike structured data, unstructured data lacks an inherently and naturally quantifiable value in a predefined way. It is more challenging to search, harder to evaluate in a consistent manner, and requires additional processing to comprehend. Most enterprise data is unstructured and is more likely to be found in qualitative than quantitative data. As such, unstructured data exists in its native, or original formats, such as audio, video, or visual formats. This form of data is stored in non-relational databases or in data lakes.

Because unstructured data cannot be processed by conventional methods, additional equipment or personnel is often required to gain insights from the data. For example, techniques such as data stacking processes large volumes of data and stacks the information in groups by similar variables. Another technique is data mining; the use of mathematical analysis software such as Oracle Data Mining (ODM) to find patterns, correlations, and anomalies to predict potential outcomes (see Figure 8).

FIGURE 8: STRUCTURED VS. UNSTRUCTURED DATA

	Structured Data	Unstructured Data
Enterprise Data	20%	80%
Data Nature	• Quantitative	• Qualitative
Data Formats	• Text & Numbers (CSV[1], XML[2])	• Variation (PDF, JPEG, WMV[3], MP3)
Storage	• Data Warehouses, Relational Databases	• Data Lakes, Non-relational Databases
Analysis	• Classification, Regression, Clustering	• Data Stacking, Data Mining

[1] Comma-separated Values
[2] Extensible Markup Language
[3] Windows Media Video

Source: Gartner, AltexSoft, Quinlan & Associates research and analysis

As the two most standard ways to describe data, defining the variety of ways in which information exists is the foundation for understanding which forms of data architecture are best suited for analysis.

4. INFORMATION SYSTEM ARCHITECTURE

After understanding the different forms and types of data, it is important to understand the data architecture that is available to store them. Though data architecture can be understood as the general framework that defines a firm's data strategy, it can also refer to the storage systems used to hold data. Major data storage providers in the market include Oracle, Amazon, Google, and Microsoft.

4.1. DATA LAKES

Data lakes are repositories that store massive quantities of non-relational, unstructured, semi-structured, structured, and log data. Consequently, data lakes require large storage capacities to be able to support all types of data, regardless of source and structure. Because of this flexibility, data lakes are also a convenient way to store unused data for later processing.

A constant influx of raw data translates to undetermined use cases for these data lakes. Without this purpose, data lakes have less organisation and filtration systems in place than data warehouses. Though this can prove difficult to interpret for those with less expertise – say business generalists – data lakes also tend to be more accessible due to their lack of structure i.e., making changes or finding information can be completed at a faster rate. This benefit is contingent on the experience of the user of the data lake; thus, analysis should primarily be conducted by data scientists who have specialised tools to translate the information for use.

The main use cases for data lakes include stream processing, machine learning, and real-time analysis. For example, IoT data analytics are stored in data lakes because the historic, aggregate IoT data is useful in identifying long-term trends and training machine-learning models at a cost-effective price. In addition to analysis, data lakes can also act as a preparation area for data warehousing. When needed, the aggregate data from data lakes can be fed into data warehouses for further processing.

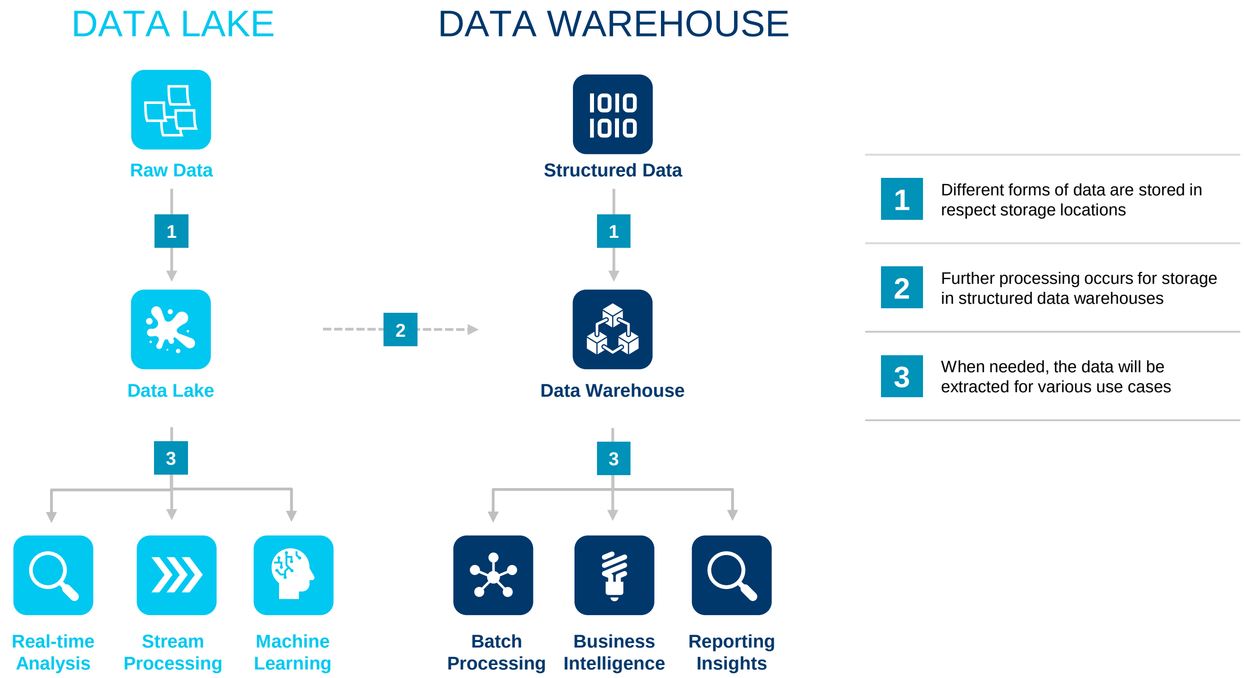
4.2. DATA WAREHOUSES

Data warehouses (DWHs) are central management systems primarily used to store refined, structured data. Given that the data has already been processed, DWHs contain less data than data lakes and store historic and relational data, avoiding expensive large capacity storage spaces by only maintaining cleansed data. DWHs normally include on-premises IT resources that use Extract, Transform, Load (ETL) to migrate data to the destination. Processed data typically have specific purposes, so the data in these repositories are currently being used to inform business decisions. As the data is stored in a familiar, structured format, it is more likely to be used by executives than data scientists.

DWH architecture may be more easily decipherable than that of data lakes, but its limits include a lack of flexibility. The architecture of DWHs tends to be more complex and rigid, meaning a simple change may take months to complete. And given the increased demand for real-time reporting and volume of everchanging data, the delayed performance of DWHs can be a major disadvantage.

The main use cases for DWHs include batch processing, business intelligence, and reporting. For example, DWHs can predict future the performance of products through data mining, which helps gain new insights (including visualisations) for business intelligence purposes. Its capability for batch processing also makes data warehouses valuable in conducting market research that requires in-depth analysis of large volumes of data.

FIGURE 9: DATA WAREHOUSES & DATA LAKES



Source: Quinlan & Associates research and analysis

5. DATA TECHNOLOGIES

In response to rapid growth in the generation and utilisation of data, various technological solutions have been developed and implemented by companies across the globe.

The four most popular technological areas include: (1) blockchain; (2) cloud computing; (3) robotics; and (4) artificial intelligence ("A.I.") (see Figure 10). These four technological areas are driving the fourth industrial revolution and are widely adopted by companies as part of their innovation / digital transformation initiatives. Blockchain was primarily developed for data storage, while cloud computing spans across data storage and application. Robotics and A.I. are typically implemented for data application purposes.

FIGURE 10: DATA TECHNOLOGIES

	DATA STORAGE		DATA APPLICATION	
	1	2	3	4
	BLOCKCHAIN	**CLOUD**	**ROBOTICS**	**A.I.**
DESCRIPTION	Distributed ledger technology that spreads the ledger across a peer network, ensuring each user has a copy of the complete ledger, only editable through consensus	Access to system resources, (e.g. storage space computing power), via the internet, providing users with scalable technology systems and enhanced functions	Hardware and / or software that can be configured to replicate human work, both physically and digitally, to automate repetitive and non-value-adding tasks	Algorithms that focus on creating intelligent machines, programmed to think, process, and evaluate information similarly to the human brain and learning mechanism
BENEFITS	Provides a single source of truth shared among stakeholders / users, ensuring data reliability	Offers flexible access to more powerful system resources, enhancing overall efficiency	Replaces unskilled labour with robotics, allowing humans to focus on value-adding tasks	Generates in-depth insights that are driven by analysis on massive amounts of data
EXAMPLES	• Client Data Management • Regulatory Technology	• Real-time Collaboration • Dynamic Pricing	• RPA • Robo-Advisor	• Predictive Analysis • Cognitive Processing
RELATIONSHIP WITH DATA	• Blockchain ensures data integrity through a unique verification process that creates robust, reliable, and traceable data, while enhancing cybersecurity	• Cloud technology helps organisations establish data infrastructures in an easy and cost-effective manner, enabling optimal processes for data usage	• Data is needed to kickstart robotic processes, and new proprietary strands of data are constantly created by the software for future procedure refinement	• A.I. software requires massive volume and a variety of data to run the algorithm and be trained for practical use, and in turn generates data for future

Source: Quinlan & Associates research and analysis

5.1. DATA STORAGE

Back-end data storage infrastructures have been developed to support the storage of data created / collected as businesses increasingly pursue data-driven operations.

5.1.1. BLOCKCHAIN

Traditionally, data has been stored and managed privately by a centralised entity to ensure security and consistency of data. However, historical incidences have shown that keeping data in a centralised manner has several shortfalls, such as exposures to cybersecurity risk, disaster risk, and internal operational risk. In addition, while data may be consistent for internal use, most organisations have different policies and protocols for data management, resulting in conflicts when sharing data during cross-entity collaborations.

To address such pain points, distributed or decentralised ledger technology can be adopted. Blockchain leverages ledgers across a peer network, ensuring each user / stakeholder has a copy of the complete ledger, which can be edited only through a consensus algorithm.

The key benefit of blockchain technology is that it provides a single source of truth, shared among different stakeholders, ensuring consistency and reliability of the data. In addition, as data on a blockchain cannot be changed without the agreement of at least 50% of network stakeholders,[4] it provides significant

[4] This depends on the blockchain protocol being used by the network.

cybersecurity against internal and external threats.

5.1.2. CLOUD

The suite of cloud offerings act as virtual data storage solutions. Instead of managing brick-and-mortar data warehouses that require physical space and operational resources, companies can leverage cloud storage services. Cloud storage services offer relatively cheap scalable data storage solutions, enabling companies to save data in a remote, secure location, and to access them via the internet.

Additionally, cloud technology is not only a storage solution, but it also involves a suite of offerings that enables users to access additional system resources for data application purposes.

5.2. DATA APPLICATION

Data application technologies are software solutions that feed on data to enhance business operations, either by replacing repetitive and mundane human tasks or producing meaningful insights by using processing power that far exceeds human processing capabilities.

5.2.1. CLOUD

Cloud solutions also offer remote access to system resources, including central processing units ("CPU"), graphics processing units ("GPU"), and random-access memory ("RAM"). This enables cost-efficient and flexible access to powerful system resources that supercharge operational efficiency.

For example, Amazon Web Service ("AWS") is a leading cloud service provider that offers a scalable platform to enable customers to deploy applications and analyse data quickly and securely. The service platform offers computing power, database storage, and content delivery solutions to help clients grow their businesses.

Combining both data storage and data application benefits, cloud technology supports organisations in establishing their data infrastructure in a cost-effective manner, enabling optimal operations for their data needs over time.

5.2.2. ROBOTICS

Robotics include physical robots and computer software. Robotics solutions execute repetitive and menial actions which traditionally required human labour, enabling employees to focus on more value-adding tasks that require creativity and critical thinking.

Many corporates adopted robotic process automation ("RPA") software for such purpose. RPA software configuration starts with feeding historical data, such as type of input, procedure, and expected output. After the implementation, the software generates additional proprietary data to help further refine the automation process. Examples of tasks RPAs can conduct include handling email responses, responding to ad-hoc requests, and cross-checking information for audit-purposes.

One RPA solution provider is UI Path, which has helped a global investment bank automate trade settlement data. The software identifies unmatched and pending trades, updates transaction numbers, and settles the trade without any human intervention. The automation has allowed banks to reduce the entire processing time from 40 minutes to just 3 minutes.

Another example of an RPA solution are robo-advisors, which can be found in many different industries. In financial services, for example, virtual robots are usually implemented on websites to start collecting data from clients regarding their background, financial position, risk appetite, investment horizon, and return goals in order to provide meaningful investment advice. The key benefits of robo-advisors are: (1) that they are much cheaper to train and operate than human advisors; and (2) they are available 24/7.

Though commonly referenced as having similar capabilities, there is a large difference between robotics and artificial intelligence: whether the software is making its own information-based decisions. Where robotics involves repetitive, pre-coded, and static tasks, artificial intelligence develops decision-making tasks that are dynamic and pre-coded for deep learning.

5.2.3. ARTIFICIAL INTELLIGENCE

Artificial Intelligence ("A.I.") represents algorithms that focus on creating an intelligent machine that is programmed to think, process, and evaluate information similarly to the human brain and its learning mechanism. With the help of A.I. and processing power, in-depth and accurate insights could be generated from big data, a previously impossible feat due to limited human capabilities.

A.I. is increasingly being used in predictive analysis, contributing and value-adding in many different scenarios, including the prevention of fraudulent activities by detecting suspicious behaviours in advance, as well as optimising marketing campaigns by predicting the right message for each identified customer segment.

One market-leading predictive modelling company is Data Robot, an A.I. company that serves organisations across a range of different industries, providing easy-to-deploy algorithms to generate accurate insights from existing data assets. For example, Data Robot provides analytics on market performance and results, enabling clients to accurately predict future demand and sales trends.

6. CONCLUSION

After understanding data forms and types and utilising the appropriate supporting architectures and data technologies, firms can begin to capture the data they need. These gold nuggets of data can begin their new journey to become valuable corporate assets down the data value chain.

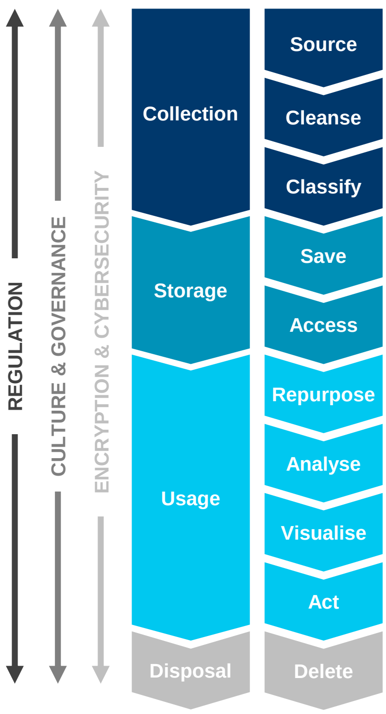
SECTION 3
DATA VALUE CHAIN

The development of a robust data strategy begins by first understanding the data value chain.

The data chain breaks down how data moves through an organisation: from collection, storage, and usage to generate insights, to disposal once the data has served its purpose.

Each step has its own procedures and issues, which impact the efficiency, accuracy, and usefulness of data-driven outputs. This chain is unified by overarching regulatory framework, a culture of data and secured by cybersecurity measures.

FIGURE 11: DATA VALUE CHAIN



- Identify the origin of data, either via creation / input by internal staff or from an external source
- Ensure each data point is accurate and standardise format across data points of the same nature
- Separate data points into relevant categories, for structured storage and to facilitate future access
- Keep the data at specified servers, which should be protected with relevant protocols
- Access the data, and transfer data from server to employee computer(s) to be used
- Re-format / Prepare data to be used as inputs, based on the protocol required by the analysis algorithm
- Evaluate the inputs (which is typically automatically completed by computer algorithms)
- Present the output in an intuitive manner, for easy understanding by the audience
- Select the appropriate action based on understanding of the visualised insights
- Remove data from storage safely, once it is confirmed that data will no longer be needed

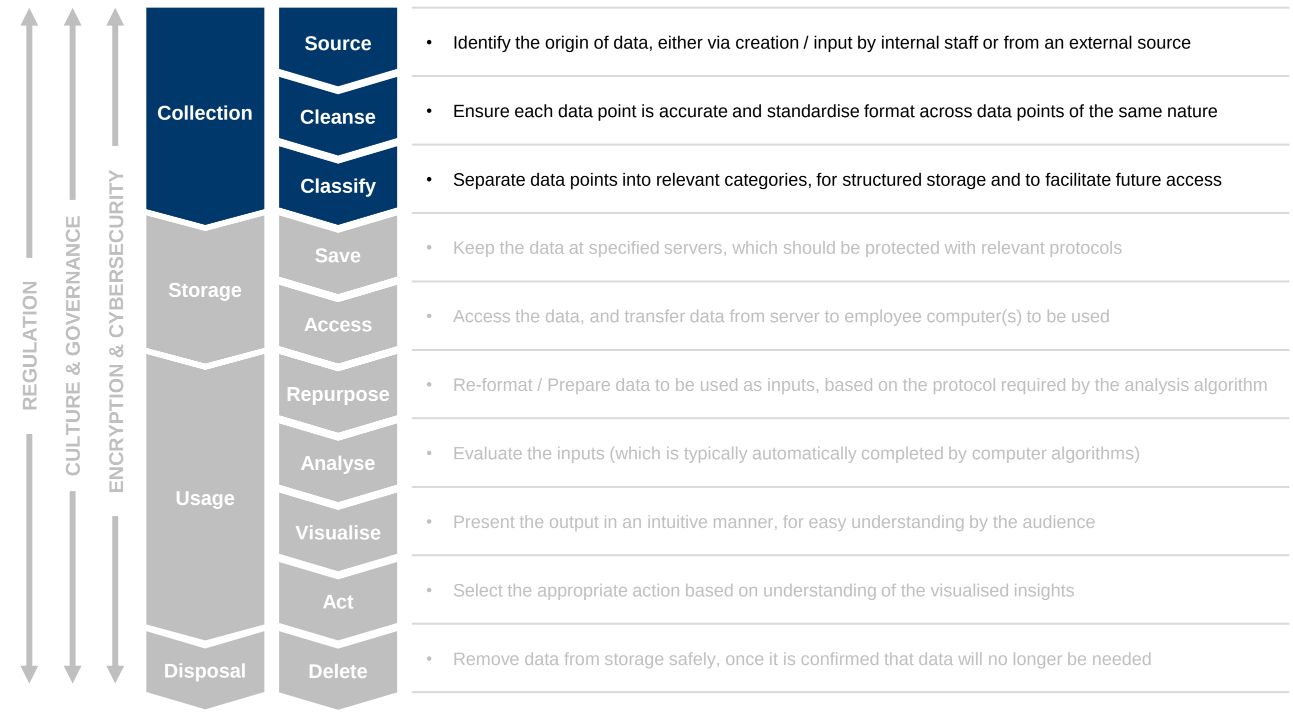
Source: Quinlan & Associates proprietary framework

1. DATA COLLECTION

The data value chain begins with the collection of data, whether from disparate sources, local sources, or from third party sources; data must be collected in the correct manner.

Poor-quality data inputs not only confuse users, but can also generate misleading or inaccurate insights, in a typical "Garbage in, garbage out" manner. Therefore, identifying the source of data and having a robust cleansing and classification procedure ensures that data is both manageable and useful for subsequent storage and use cases.

FIGURE 12: DATA VALUE CHAIN – COLLECTION



Source: Quinlan & Associates proprietary framework

1.1. SOURCE

Data can be sourced in various ways, including both internal and external channels. The origin of the data is of tremendous importance, as it primarily determines whether the data is structured or unstructured, affecting its usage at later stages of the value chain.

Key considerations for this stage include defining what the data is being used for and what target insights are desired, alongside the relevance and accuracy of the data collected. Taking these considerations into account early in the value chain not only helps companies in filtering important data from noise, but it also ensures that they comply with data regulations specific to the region or industry they operate in. This will be further elaborated in the

'Regulation' section of the value chain later in the report.

1.2. CLEANSE

After sourcing, companies should cleanse the collected data – unstructured data should be converted into structured data via a standard methodology, datasets should be reviewed based on coverage, and any overlaps should be eliminated. Additionally, while reviewing data coverage, any gaps should be regularly filled via an expansion of data procurement.

1.3. CLASSIFY

Once cleansed, data should be sorted and labelled based on pre-defined categories, then transferred to the relevant database for storage and future access.

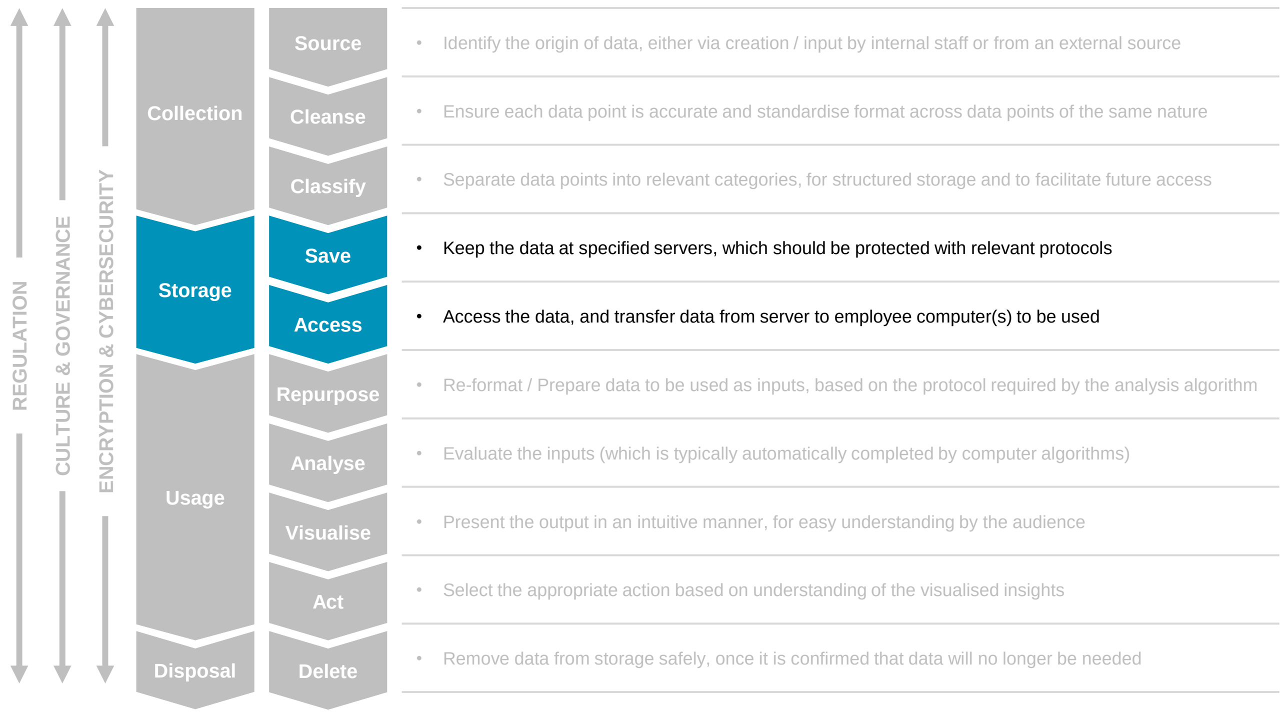
This requires companies to establish their own data classification system, typically based on their industry, operations, and practices. While there are industry-wide terminologies, these indices should be tailored and customised to reflect a company's operations and anticipated usage for the data. A structured and holistic indexing system is fundamental to ensure consistency and accessibility of data across the organisation.

A STRUCTURED AND HOLISTIC INDEXING SYSTEM IS FUNDAMENTAL TO ENSURE CONSISTENCY AND ACCESSIBILITY OF DATA ACROSS THE ORGANISATION

2. DATA STORAGE

While data storage does not directly affect the use of data, it still represents a critical stage in the data value chain. Data must be maintained and protected for a company to function properly and sustainably. Data protection and encryption is further elaborated in the 'Encryption & Cybersecurity' section, later in the report.

FIGURE 13: DATA VALUE CHAIN – STORAGE



Source: Quinlan & Associates proprietary framework

2.1. SAVE

Once data is properly categorised, it is ready to be saved in its designated database. Traditionally, companies purchased and operated physical servers (rack space) and hired relevant staff to protect and maintain the servers. However, with the development of communications technology and cloud solutions, many businesses are adopting virtual databases. As discussed in our previous report, Banking on the Cloud,[5] benefits of cloud technology in data storage include scalability, reliability, and cost-efficiency.

When planning or shifting data across databases, companies should be mindful of regulatory standards set forth by local regulators in countries of operations. For example, many countries have imposed data localisation rules that restrict sensitive information from leaving the country, which typically affects companies with global operations.

[5] See Quinlan & Associates, Banking on the Cloud: Supercharging Collaboration Through Cloud Technology, available at: https://www.quinlanandassociates.com/insights-banking-on-the-cloud/

Saving data across jurisdictions presents a key challenge when it needs to be accessed and pooled for analysis. Should a company operate multiple databases across multiple jurisdictions, the best practice is to follow the most stringent rules and extrapolate these standards to all operations to prevent potential regulatory setbacks caused by the company's local or regional data management practices.

2.2. ACCESS

One of the key purposes of data classification is to categorise data based on the level of sensitivity in an organisation whilst aiding access control protocols. Access controls should be implemented by granting relevant access to only qualified individuals, minimising the risk of data leakage and tracking leaks down to an individual level for post-mortems by a cybersecurity team.

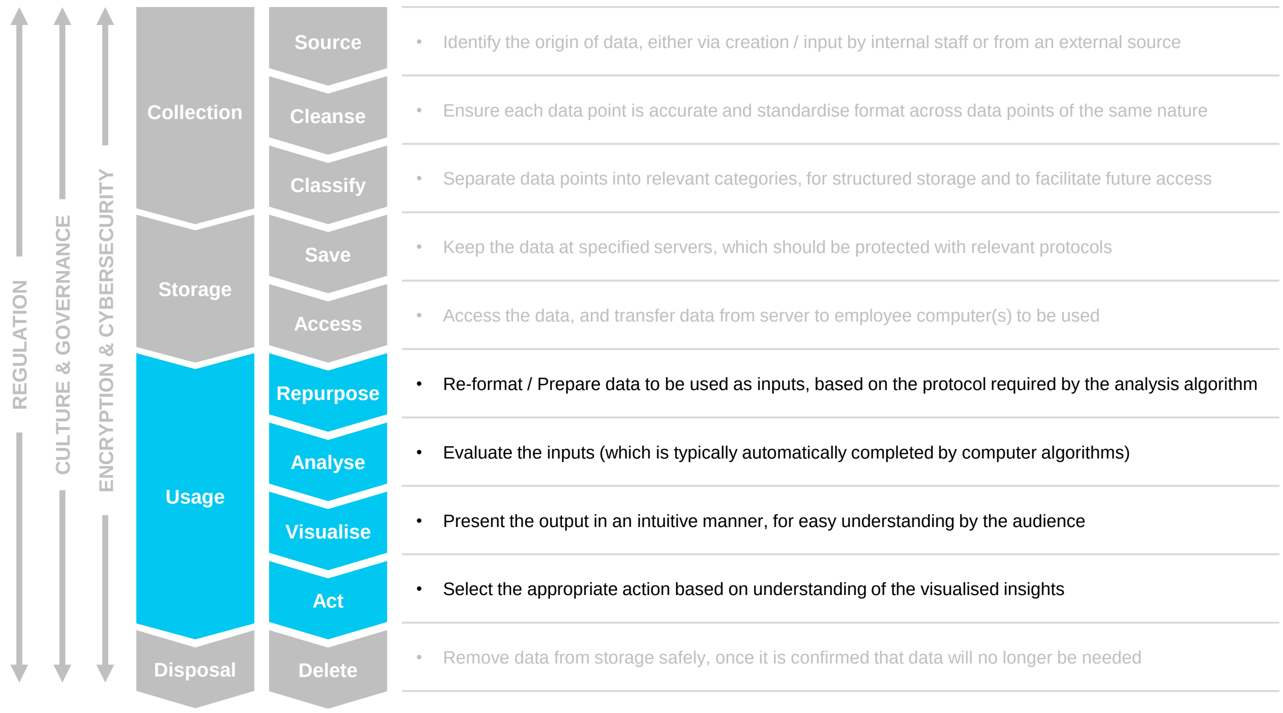
ACCESS CONTROLS SHOULD BE IMPLEMENTED BY GRANTING RELEVANT ACCESS TO ONLY QUALIFIED INDIVIDUALS, MINIMISING THE RISK OF DATA LEAKAGE

3. DATA USAGE

Using data for generating insights has garnered the most amount of attention in the data value chain over the last decade. And rightly so, given it is the portion of the value chain that generates visual elements for use and creates actionable recommendations which affect day-to-day business decisions.

However, as mentioned at the start of the report, this is merely one step in in an organisation's overall data strategy. Blind analysis of data can lead to catastrophic consequences for a business that considers themselves "data driven". Actual usage of data in a proper manner requires several steps after pulling it from its dormant database, including: (1) repurposing it or preparing it for analysis; (2) analysing it for a specific purpose; (3) visualising it to showcase insights / recommendations that are easy to interpret; and (4) acting upon the recommendations given.

FIGURE 14: DATA VALUE CHAIN – USAGE

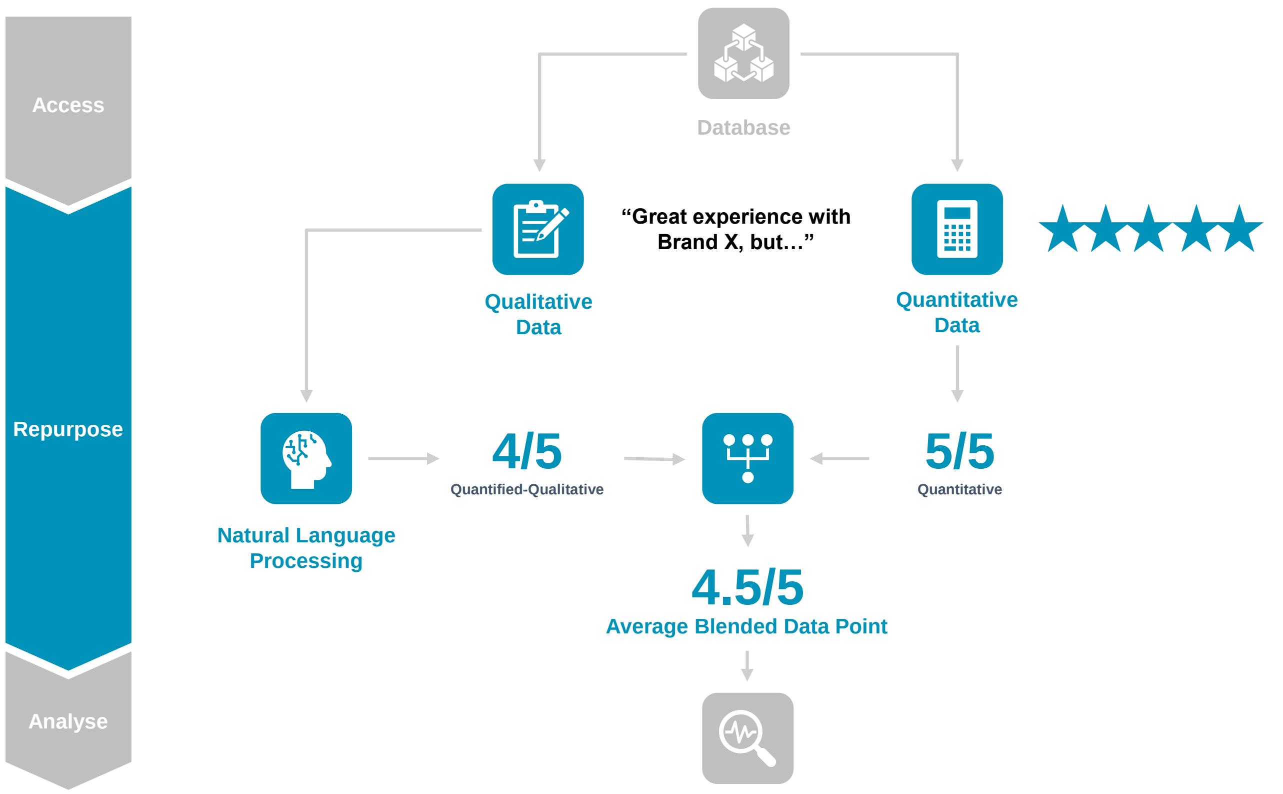


Source: Quinlan & Associates proprietary framework

3.1. REPURPOSE

Before analysing data, the user must prepare and tailor data according to the subsequent analysis methodology. An organisation may use multiple software programmes to analyse and visualise information; as a result, data from storage needs to be repurposed to ensure it is compatible with target analysis algorithms.

FIGURE 15: DATA REPURPOSING



Source: Quinlan & Associates proprietary framework

For example, there may be a dashboard that showcases customer reviews for a product over time. Reviews consists of a 5-star system and a feedback section; this is considered both quantitative data (5 stars) and qualitative (text in the feedback section). If the dashboard was looking to visualise customer sentiment towards the product over time, the qualitative data (text from feedback) would need to be analysed by a natural-language processor and assigned a quantitative value (0 to 5), converting it into "quantified-qualitative" data, which can then be averaged into a unified score for visualisation (see Figure 15).

3.2. ANALYSE

After repurposing data and categorising it accordingly by quantitative, qualitative, or quantified-qualitative, it is ready for analysis.

There are four types of data analysis, namely: (1) descriptive; (2) diagnostic; (3) predictive; and (4) prescriptive analysis (see Figure 15).

Each layer generates its own level of value to an organisation and provides a foundation to the next. Descriptive and diagnostic analysis might appear trivial, but without it, predictive and prescriptive analysis would not be possible. Prescriptive analysis being the most sought after by organisations as it not only predicts the outcomes of different scenarios, but also recommends the appropriate action for a given scenario.

As data analysis moves from the bottom to the top, it becomes increasingly difficult to produce top layer analysis as the next layer depends on near perfect execution of the former. However, if done correctly, a business gains the ability to foresee and avoid possible pitfalls, delivering it a real-world competitive advantage over its competitors.

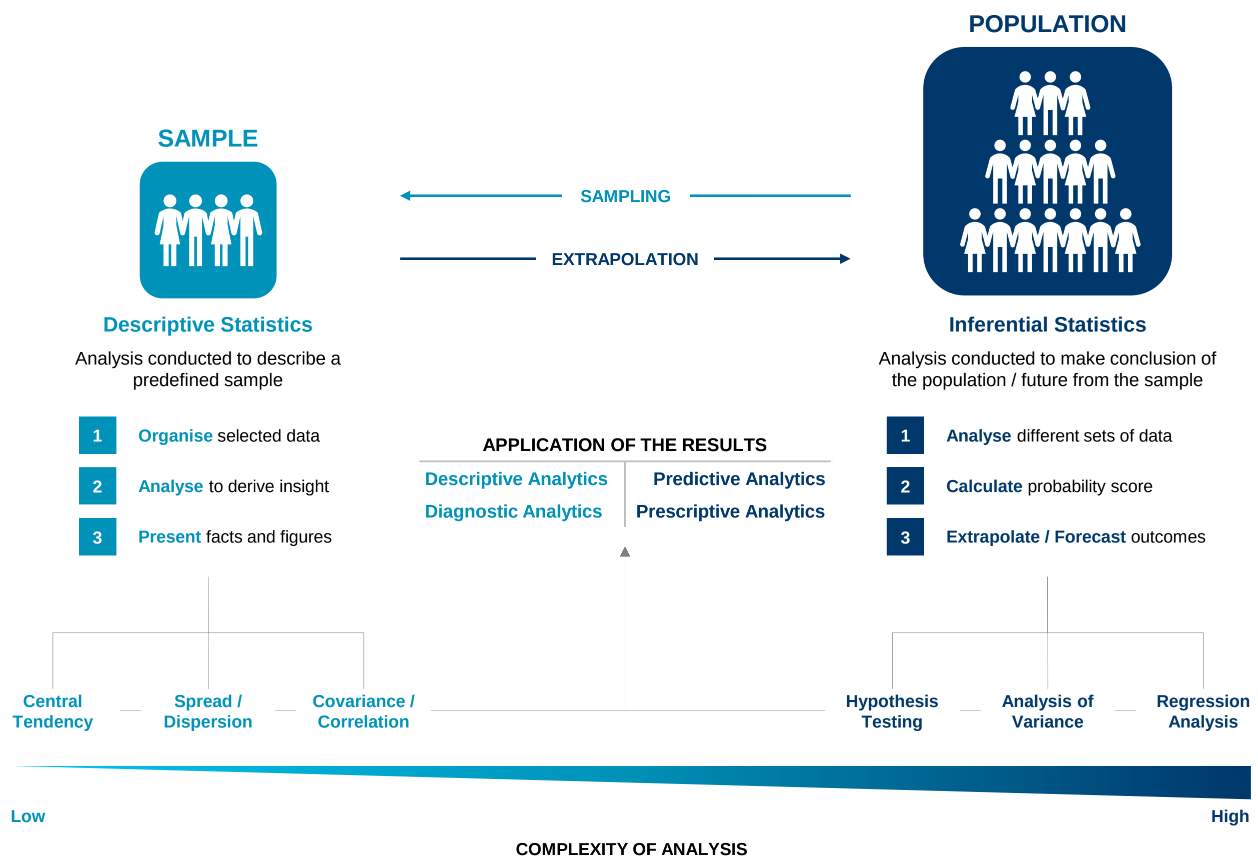
FIGURE 16: FOUR TYPES OF DATA ANALYSIS

	DESCRIPTION	USAGE	
PRESCRIPTIVE	• Provides anticipated outcome of different actions	• Recommends an optimal action for the situation	Forward-Looking
PREDICTIVE	• Calculates probabilities of future events	• Predicts the likelihood of different situations	Forward-Looking
DIAGNOSTIC	• Identifies relationships between variables	• Understands why events occurred the way they did	Backward-Looking
DESCRIPTIVE	• Transforms datasets into information	• Describes events that have already occurred	Backward-Looking

Source: Quinlan & Associates research and analysis

Descriptive statistics and inferential statistics represent the two most common statistical analysis methodologies in business. Descriptive statistics aims to describe facts and figures of a selected sample set, while inferential statistics focuses on extrapolating information to make conclusions on a larger population (see Figure 17). Both methodologies generate four types of data analysis shown above.

FIGURE 17: TYPES STATISTICAL ANALYSIS METHODOLOGIES



Source: Quinlan & Associates research and analysis

3.2.1. DESCRIPTIVE STATISTICS

Descriptive statistics are used to organise and analyse facts and figures to derive insights, which are typically demonstrated in visual form, such as charts and tables.

Popular business metrics for descriptive statistics include central tendency, spread / dispersion, and correlation / covariance. Central tendency typically refers to the average of the sample, dispersion refers to the deviation of data points from the centre, and correlation / covariance measures the relationship and co-movements between multiple data points.

Descriptive statistics can be calculated relatively easily and has the benefit of organising, summarising, and presenting data in a digestible format. As a result, descriptive statistics is applied regularly in business to produce various forms of collateral for communication. This type of statistical analysis benefits the most from quantitative data, which can be easily compared and processed by even the most rudimentary tools, including Microsoft Excel.

3.2.2. INFERENTIAL STATISTICS

While descriptive statistics describes the selected sample set, inferential statistics aims to infers insights about the wider population. This is typically conducted via calculating probability scores and extrapolating results to make informed conclusions on the population or on future events.

To extract insights, more advanced statistical tools (nonetheless, most basic statistical programmes can complete these with relative ease), such as hypothesis testing, analysis of variance ("ANOVA"), and regression analysis, are used in addition to typical tools used for descriptive statistics. In layman's terms, inferential statistics aims to: (1) understand the relationship between various factors / drivers in a predefined sample; and (2) estimate the probability of the sample being representative of the population.

As inferential statistics can be used for predictive and prescriptive analytics, many companies are actively developing solutions in inferential statistics, including A.I.-based solutions, to understand relevant trends in advance and to maximise upside. This type of statistical analysis can utilise all data formats depending on the solution provided. However, it is important to note that, like most programs, it can offer the best prescriptive or predictive prowess when paired with structured data as opposed to inferring once from qualitative data, which can be influenced by noise – for example, the two-step process of inferring that a cat exists in the foreground of an image and ignoring the rest of the background, followed by predicting whether a cat exists in a new image.

3.3. VISUALISE

Once the data has been adequately analysed, it can hold a significant amount of insight. However, it may be challenging for a user to process it in a purely table or text format. As such, visualisation represents a key step in the data value chain for ease of understanding.

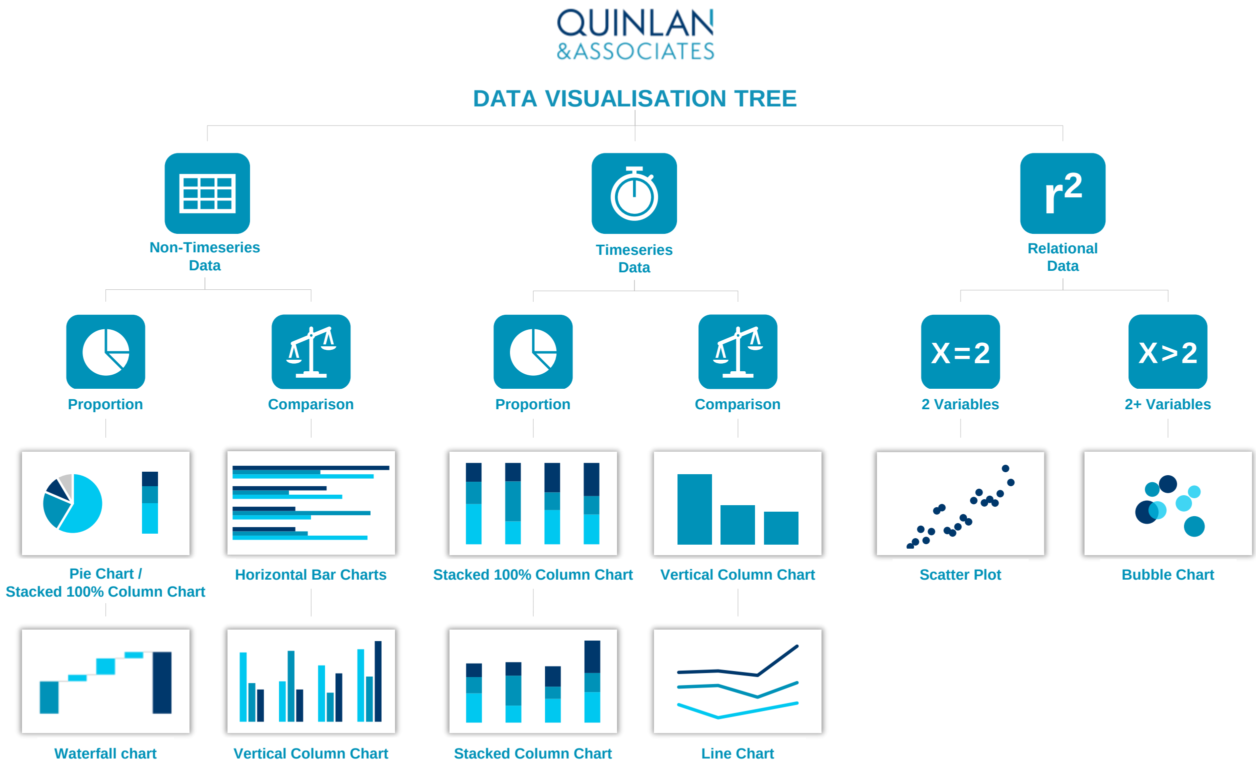
According to the MIT Department of Brain and Cognitive Science, 50% of the brain is directly or indirectly devoted to vision.[6] The human brain takes information much more easily and quickly when it is presented via a visual cue. Accordingly, a key benefit of visualisation is as an effective and scalable medium of communication. Visual representation of data is essential for any business presentation, as the audience must understand and interpret the information quickly to make informed decisions.

Many individuals struggle to visualise their analysis in an effective manner. This challenge is only exacerbated by the abundance of visualisation choices made available – in fact, in PowerPoint alone, there are 38 types of charts across 16 categories.

A decision tree can be used to help decide the type of chart to be used for visualisation. Leveraging on past experience, Quinlan & Associates has developed a proprietary decision tree for visualisation, based on the most popular types of charts (see Figure 18).

6. Massachusetts Institute of Technology, Brain Processing of Visual Information, available at: https://news.mit.edu/1996/visualprocessing

FIGURE 18: VISUALISATION DECISION TREE



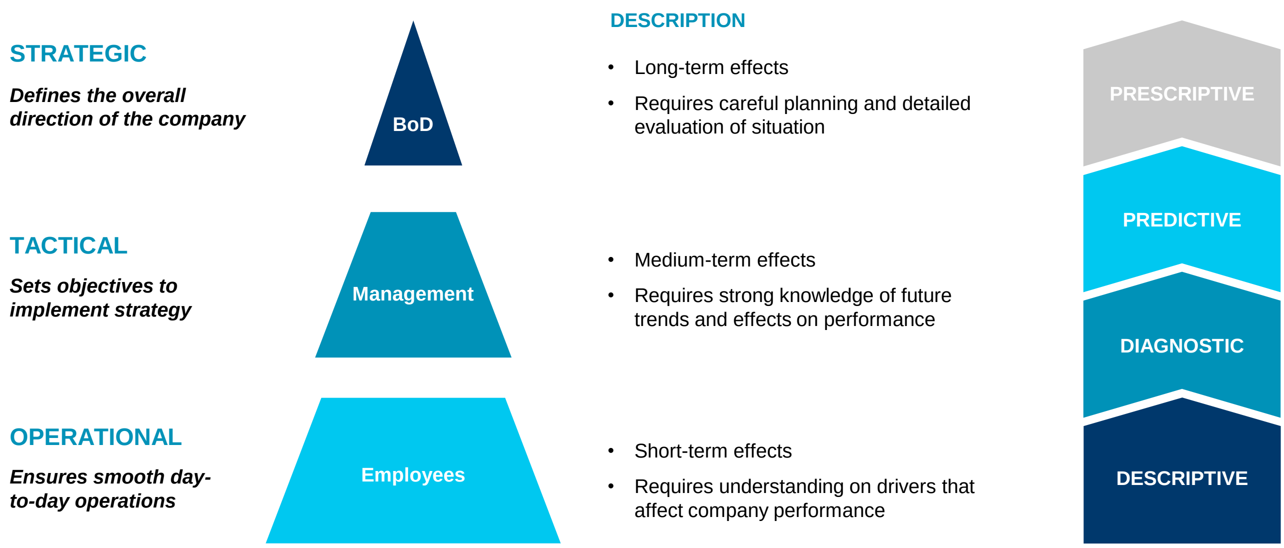
Source: Quinlan & Associates proprietary framework

Various dimensions, based on the underlying data and / or analysis conducted, should be considered in an ordered manner to determine which type of chart presents the information in the most effective manner.

3.4. ACT

After being presented output from analysis and understanding any associated recommendations, decision makers are required to agree on a suitable course of action from data. There are three main layers of actions within most organisations, based on the level and responsibility of the decision maker (see Figure 19).

FIGURE 19: LAYERS OF COMPANY ACTIONS



Source: Quinlan & Associates research and analysis

3.4.1. STRATEGIC

Strategic decisions form the underlying foundation of any company on which all other actions are based. As the most crucial decisions with significant impacts on the future of an organisation, strategic decisions require the most sophisticated form of analysis and decision-making, requiring a strong understanding of future external trends, potential actions (and reactions), and associated implications for the company. As such, prescriptive analysis is often the most appropriate solution.

Companies without the resources or technological capabilities to leverage prescriptive analysis can also opt to use predictive analysis, providing them with a comprehensive view of the future prior to generating any strategic ideas (and understanding potential results).

3.4.2. TACTICAL

Tactical decisions act as the link between strategy and execution and facilitate the implementation of an organisation's strategy. There are two key considerations for these decisions, namely: (1) the implications of future trends on industry drivers; and (2) how industry drivers affect performance. Predictive analysis can be leveraged to understand future trends, while diagnostic analysis can be used to identify and quantify the effects of those trends on a company's performance.

3.4.3. OPERATIONAL

Operational decisions affect the day-to-day running of a company, and the evaluation process is conducted frequently and by most employees. These actions typically have relatively immediate effects, and therefore predictive and prescriptive analyses may not be as relevant. Furthermore, because decisions made at this level are typically less impactful, mistakes may not be as costly. As such, the analysis used can be more elementary and, for the most part, are typically descriptive and diagnostic in nature.

STRATEGIC DECISIONS REQUIRE THE MOST SOPHISTICATED FORM OF ANALYSIS AND DECISION-MAKING, REQUIRING A STRONG UNDERSTANDING OF FUTURE EXTERNAL TRENDS, POTENTIAL ACTIONS (AND REACTIONS), AND ASSOCIATED IMPLICATIONS FOR THE COMPANY

4. DATA DISPOSAL

Unnecessary data should be deleted from systems and the data warehouse at the end the value chain, as data minimisation is essential to the reduction of risk of unauthorised access and other cybersecurity threats.

Furthermore, regulators are beginning to restrict companies' abilities to keep data perpetually due to privacy concerns, and some have imposed fines on companies failing to comply with data disposal practices. For example, Telecom Italia was fined EUR 27.8 million in early 2020 by Italian data protection authorities due to storing data beyond the time limits established by company policies (5 years) and utilising it for promotional purposes without consent.[7]

FIGURE 20: DATA VALUE CHAIN – DISPOSAL

Source: Quinlan & Associates proprietary framework

To this end, companies should regularly examine their data warehouse to identify and dispose of any data that has already served its purpose. This process can be facilitated by implementing a suitable classification protocol during the data collection stage. If a company labels each data point with an expiration date, data can be automatically discarded by the server at the right time.

7. BBC, Three years of GDPR: the biggest fines so far, available at: https://www.bbc.com/news/technology-57011639

5. ENCRYPTION & CYBERSECURITY

Cybersecurity is an area that is often overlooked in favour of data analysis outcomes. It does not draw attention of management until a public breach occurs at great financial or reputational cost. Therefore, it is vital that cybersecurity underpins the entire data value chain, securing company data from both internal and external threats of unauthorised access. Cybersecurity can be split into tools and policies used to protect the value chain.

FIGURE 21: DATA VALUE CHAIN – CYBERSECURITY

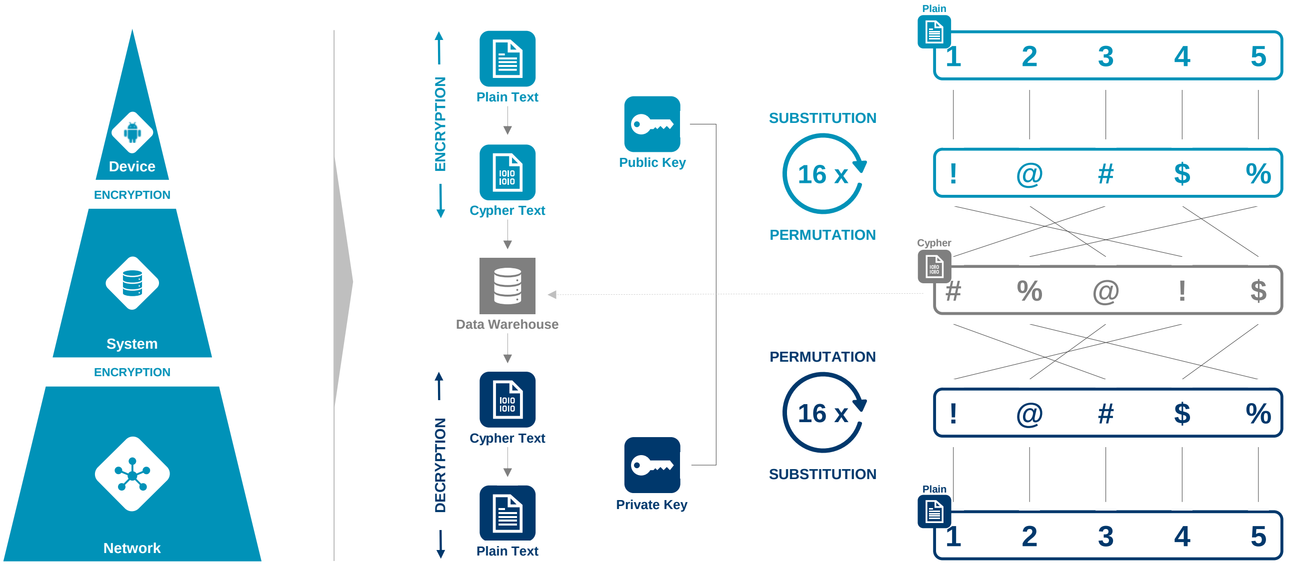
Source: Quinlan & Associates proprietary framework

5.1. CYBERSECURITY TOOLS

Any information or data sourced by companies enters the ecosystem in a form of plain text, which can be accessed, viewed, and shared by anyone. Access control represents the first level of defence against data leakage. All data-at-rest (i.e. data not being actively used) should be fully encrypted before being transferred to the data warehouse, for an extra layer of data protection.

Encryption is the next layer of protection, preventing anyone from unsolicited viewing (see Figure 22). This step typically occurs after collection of data and before data storage. However, in reality, the entire value chain should remain encrypted until a specified use case presents itself, such as the permanent disposal of data from an organisation or for an autopsy in the event of a security breach.

FIGURE 22: DATA ENCRYPTION



Source: IBM, Quinlan & Associates analysis

This encryption method involves two steps: (1) substitution; and (2) permutation.

Substitution converts each letter of the plain text data into a different letter based on a pre-defined conversion table, masking the original information. This step is also known as "confusion", as the output from this process aims to confuse unsolicited readers.

Permutation randomly re-arranges the order of the substituted letters, adding randomness into the encrypted information. The random re-arrangement of letters provides an extra layer of confusion, while randomness makes the decryption without the decryption key virtually impossible.

Substitution and permutation are typically conducted 16 times to minimise the chance of breaking through the encryption without the decryption key. A decryption algorithm, along with the decryption key, should be provided to authorised users to uncover the original data.

Apart from encryption, network security via firewalls, anti-malware solutions, and managed security solutions that monitor and implement incident responses, are critical tools in securing data from undesirable parties.

5.2. CYBERSECURITY POLICIES

Cybersecurity is not only about the tools at a company's disposal to thwart external attackers from a company's trove of data. Attacks come in all shapes and sizes. In fact, 75% of data breaches are attributed towards human factors or human error.[8] Hence, a company should implement appropriate policies and measures which look to employees as the first line of defence against breaches.
Companies need to provide frequent and robust education (such as via workshops and

8. NCBI, Human factor, a critical weak point in the information security of an organization's Internet of things, available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7980069/

conferences) to all levels of employees to ensure understanding of the risks associated with cybersecurity. Companies also need to properly educate staff members on all cybersecurity software and authentication applications implemented, including how they function and why they are needed for safe business operations. At the system- and network-levels, companies should hire an IT team that fully understands the organisations' IT infrastructure to provide proper oversight over information security. Assigning proper authorisation and access rights to only relevant personnel is crucial to protect sensitive and confidential data.

IT policies and procedures should be effectively communicated to staff with regards to data privacy and risk control standards. Standard operating procedures should be reviewed regularly, with the frequency of updates depending on the level of information sensitivity and when regulatory changes occur.

Finally, should a breach occur, a comprehensive contingency plan should be in place to escalate the issue quickly and directly to the correct stakeholders at the management level. This is so that isolation and damage control can be implemented effectively. To this end, efficient internal and external communication channels are required. All breaches should be documented carefully to enhance measures and systems to prevent similar breaches in the future.

Comprehensive data protection across the data value chain requires a high level of coordination across the company, which is dependent upon well-designed resilience tests, comprehensive recovery plans, and regular staff training. A robust cybersecurity posture cannot exist without the policies and culture necessary to support it. To understand the cybersecurity requirements needed, then, it is important to review the regulations surrounding data and how the organisation guides its employees in its usage on a day-to-day basis

75% OF DATA BREACHES ARE ATTRIBUTED TOWARDS HUMAN FACTORS OR HUMAN ERROR

6. CULTURE & GOVERNANCE

The data value chain is, at its core, an ideal process companies can look to follow when handling data. However, organisations, at the end of the day, are made of people. Whether deliberate or unintentional, individuals tend to take the path of least resistance when it comes to work. If that means creating a temporary shortcut which affects the data process downstream, it will create more work for a business to recover and operate efficiently again.

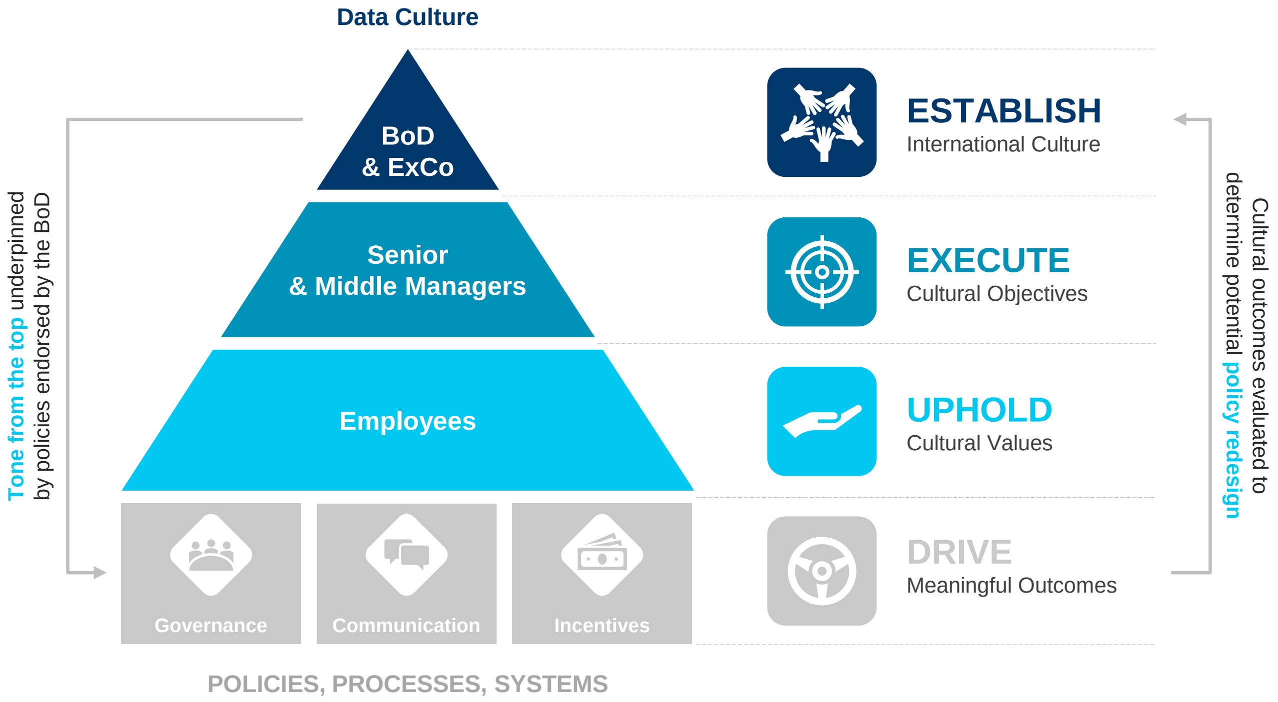
In reality, the human factor in data handling is often overlooked in favour of the tools used, but without a clear and concise culture around data within an organisation, we believe that many of the tools and processes deployed simply fall by the wayside.

FIGURE 23: DATA VALUE CHAIN – CULTURE & GOVERNANCE

Source: Quinlan & Associates research and analysis

As highlighted in several of our previous reports, culture is anchored by three main pillars, namely: (1) Governance, (2) Communication, and (3) Incentives, all of which are supported by relevant policies, processes, and/or systems to drive delivery. Once the culture is set, it becomes a self-sustaining feedback loop, where cultural outcomes from a positive data culture begins to shape the organisation as a whole from the top to the bottom (and vice versa).

FIGURE 24: DATA CULTURE FRAMEWORK



Source: Quinlan & Associates research and analysis

6.1. GOVERNANCE

The first step in establishing a data centric culture begins with governance policies set from the top of an organisation, including the board of directors and executive committee.

The upper echelons of an organisation should produce guiding policies around the use of technology systems which collect, store, and use data in a business setting. These policies set the tone for mid-level managers, which then cascades down throughout a company to build company-wide standards.

An example of where data governance often goes wrong is in the use of Customer Relationship Management ("CRM") systems. Top sales executives typically communicate over email and instant messaging tools, with many leads not transferred into a company's CRM system. Moreover, siloed, non-standardised data is often added to company CRMs, creating confusion and reinforcing suboptimal behaviours.

It is not hyperbole to say that the devil is in the detail for everyday processes and policies that influence everyone in an organisation. In keeping with the CRM example, a simple policy surrounding lead generation, which would include the inputting of data into a CRM for subsequent follow ups, would have been the first step in the governance process.

In our upcoming report, we not only demonstrate how policies help shape the data culture of an organisation, but also how it promotes lasting change in times of organisational transformation.

6.2. COMMUNICATION

Policies alone will not shape a culture; there must be a process in place to communicate these guidelines and instructions across the entire company hierarchy.

Communication is a key foundation needed to execute upon data policies put in place and to meet cultural objectives. Once policies are approved, direct lines of communication channels to middle managers and ground level employees should be in place to disseminate the relevant information to appropriate parties. More importantly, employees often have the best understanding of the effectiveness of policies at the ground level, making it highly beneficial to have communication channels go from the bottom up as well.

Allowing for employee direct input creates buy-in into an organisation's vision of a data centric culture. This feedback loop is critical for a business to stay on top of changing systems and provide the best processes for relevant employee. Looking again at the CRM example, if a company was migrating from one CRM solution to another, employees may notice that policies and processes that were once used may not be effective anymore and would need a way to relay this information upwards to drive the changes needed. This promotes a positive culture of adaptability, highly desirable in a digital environment where efficiency and effectiveness are prized.

6.3. INCENTIVES

Companies must institute a robust incentive system to reinforce policies and processes that have been implemented.

Human beings are creatures of habit, and breaking unhealthy habits (in this case, an old business process) requires significant amounts of repetition and the appropriate motivation to do so. Companies can help this along by providing a system of incentives for desired process and policy adherence. For example, using the CRM example again, a company could track the sale of a product to the original employee who inputted the lead into the CRM, rewarding him or her appropriately for their effort while creating a system of adherence that reverberates across all business lines. However, it is all-too-common for many firms to use pure input based KPIs (e.g. "you must log 20 call reports per month"), which frequently results in a garbage-in, garbage-out outcome for many companies around the world.

In summary, the culture and governance of an organisation spans across the data value chain. Without appropriate data governance policies, communication protocols, or incentive systems in place, the data value chain does not function as desired. To craft these policies, processes, or systems holistically, one should begin by reviewing the regulations surrounding data, which often shapes the initial outlook on how data use is both facilitated and monitored by employees within an organisation.

7. REGULATION

Many regulatory bodies have already developed and implemented data laws based on country-specific circumstances, and mismanagement of data may result in regulatory consequences.

As issues around data privacy and ownership increasingly become a concern, regulation must also be a consideration throughout the entire data value chain (i.e. from sourcing to disposal), and companies must develop an in-depth understanding of the implications and monitor changes to adapt data their strategy and operations accordingly.

FIGURE 25: DATA VALUE CHAIN – REGULATION

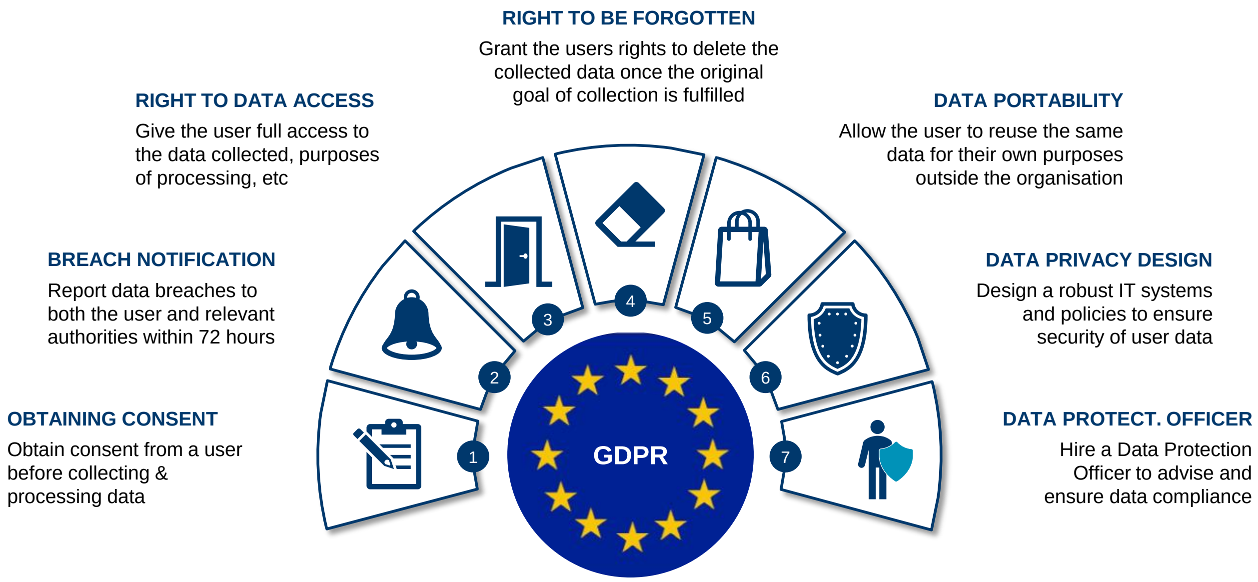
Source: Quinlan & Associates proprietary framework

7.1. GENERAL REGULATIONS

Implemented in May 2018, the General Data Protection Regulations ("GDPR") is one of the most widely applicable frameworks for collecting and processing data on residents from the European Union ("EU"). The key purpose of the GDPR is to designate customers as the ultimate data owner and to increase transparency throughout a company's data value chain. In addition to providing data management guidelines across the value chain, the GDPR framework also establishes a standard for data protection and a list of compliance requirements.

The GDPR governs policies across seven areas: (1) obtaining consent; (2) breach notifications; (3) right to data access; (4) right to be forgotten; (5) data portability; (6) data privacy design; and (7) data protection officers (see Figure 26).

FIGURE 26: GENERAL DATA PROTECTION REGULATION



Source: European Parliament and Council of the European Union, Quinlan & Associates analysis

7.1.1. OBTAINING CONSENT

GDPR requires organisations to seek consent from the data provider after clearly stating the purpose for and implications of data collection and before collecting and storing any personal or sensitive data. The regulation further empowers data providers by enabling them to withdraw consent at any time.

7.1.2. BREACH NOTIFICATION

A data breach is a security incident that affects the confidentiality, authenticity, or availability of internally generated or externally collected data. A breach can be internally-driven (such as an employee accidentally sending sensitive information to the wrong person) or externally-driven (such as a hacker successfully penetrating the system and accessing personal data).
Companies are required to notify users, data controllers, and relevant supervisory authorities of any data breaches within 72 hours of identification. The notification should include the nature and the number of users affected by the data breach, and the measures to be implemented by the organisation. Failure to report breaches within the 72-hour timeframe will result in fines.

7.1.3. RIGHT TO DATA ACCESS

Upon request, the company must provide, free of charge, a detailed electronic copy of all data collected on the user, the ultimate data owner. The report must state the types of data collected, purpose of collection, parties involved, and destination countries if data has been transferred (if applicable).

7.1.4. RIGHT TO BE FORGOTTEN

Users have the right to request one's data to be deleted when: (1) the original purpose of data collection has been fulfilled; (2) the user withdraws consent; or (3) personal data has been unlawfully processed. This pillar supports a key goal of GDPR, data minimisation, which stipulates that an organisation must carry out businesses with the least amount of data necessary.

7.1.5. DATA PORTABILITY

Users have the right to obtain one's own data from the company and to reuse the data for personal use outside the company. Upon request, the data needs to be sent to the user in a structured and interoperable format (e.g., machine-readable format).

7.1.6. DATA PRIVACY DESIGN

Organisations are required to implement appropriate procedural and technical measures to ensure privacy. To this end, organisations should develop firm-wide policies, processes, and protocols to drive data protection and should maintain robust IT systems to ensure cybersecurity proficiency.

7.1.7. DATA PROTECTION OFFICER

A company must designate a data protection officer ("DPO") if the company is a public authority / body or if the company's activities involve large-scale collection of personal data. The DPO's responsibilities include ensuring internal compliance, supervising the data protection obligations of the company, and serving as the key contact point for users and the supervisory authority for data usage-related enquiries.

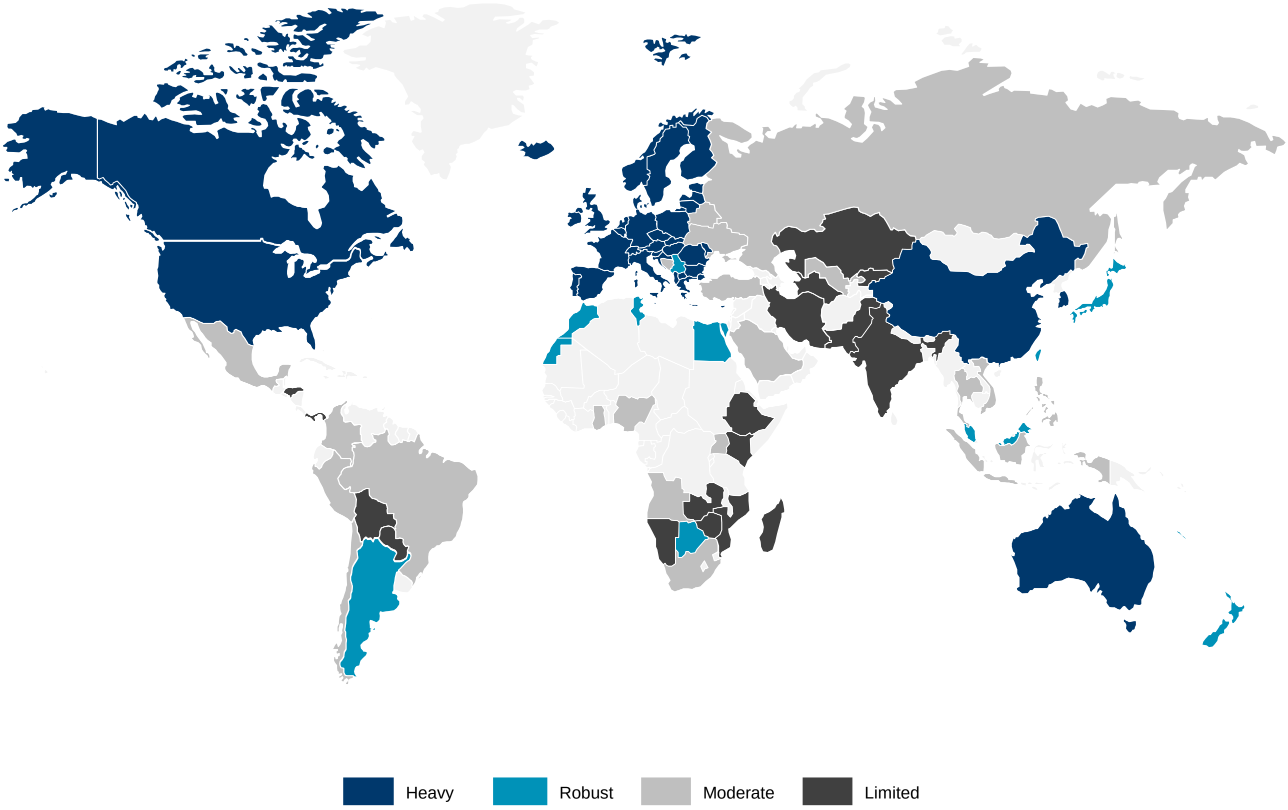
7.1.8. GDPR APPLICABILITY

The GDPR is only enforced in member states of the EU, and hence, is relevant for businesses in the region or conducted online. Furthermore, opting to follow more stringent regulations is recommended for global organisations when choosing amongst various local data requirements. As such, even though GDPR may not be directly applicable to companies outside of the EU, they should view GDPR as a minimum standard when considering their own data value chain, especially if they currently – or intend to – operate internationally.

7.2. LOCAL REGULATIONS

In addition to the GDPR, most developed countries have already implemented robust data laws based on the jurisdiction's specific priorities (see Figure 27). The key themes evolve around data privacy, data localisation, and incorporation of cybersecurity. Data privacy empowers and protects the rights of users, data localisation requires certain data of citizens to stay within the jurisdiction (to prevent information from being abused outside the territory), and cybersecurity maintains robust internal systems to protect information from external cyberattacks.

FIGURE 27: ROBUSTNESS OF DATA PROTECTION LAWS AROUND THE WORLD



Source: : DLA Piper, Quinlan & Associates analysis

In North America, both the U.S. and Canada have implemented their own stringent data laws, centred around data privacy and cybersecurity. While data localisation has been discussed in the U.S. since 2013 for counter-terrorism purpose, no concrete laws or requirements have been established. On the other hand, certain provinces in Canada, including Nova Scotia and British Columbia, have implemented data localisation policies.

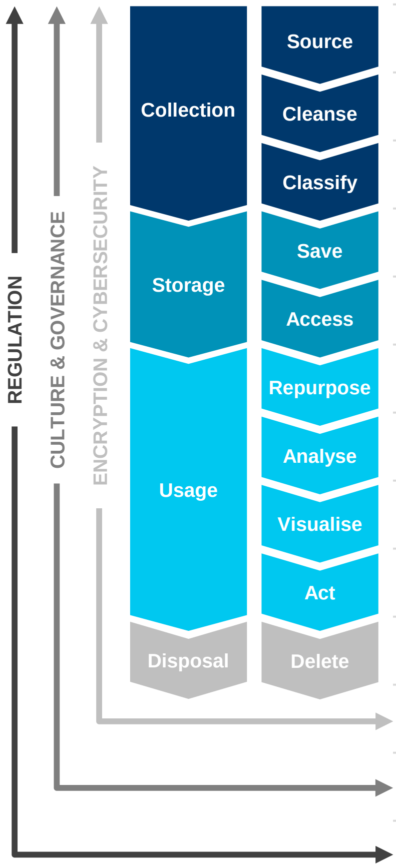
More developed countries in the Asia Pacific ("APAC") region, such as China, Japan, South Korea, Australia, and New Zealand, have established laws that focus on data privacy and cybersecurity. While not as widespread, data localisation has also been a key focus for many countries. For example, China requires all personal, business, and financial data to stay within the country. Other countries have imposed data localisation requirements on specific datasets, such as Australia with respect to health records and South Korea on map data (for national security reasons).

While data laws differ across jurisdictions, companies must understand the implications of local data requirements to refine their operations across the data value chain and build a robust data strategy.

SECTION 4
CONCLUSION

As highlighted in this paper, there are a number of common issues that plague many companies across each step of the data value chain, ranging from the seemingly minute details of how you save a file to the highly visible phishing scams which seek to breach an organisation (see Figure 28). A complete data value chain would include solutions which match an organisation's business needs. We believe only 5% of companies globally, excluding micro unlisted firms and NGOs, have adequately fulfilled the entire value chain.

FIGURE 28: SUMMARY OF COMMON ISSUES & SOLUTIONS ON THE DATA VALUE CHAIN



Common Problems	Ideal Solutions
• Disparate / ambiguous data sources from both internal and 3rd party data systems	• Complete information mapping / taxonomy, coupled with 3rd party data vendor policies
• Absence of or incomplete / outdated data cleansing processes	• Unified data cleansing processes based on business domain
• Inadequate data classifications based on industry, operations, or practices	• Complete and accessible data catalogue / index for employees to follow
• Inconsistent saving protocols and/or file paths for data points	• Combination of automated saving protocols and employee training on data saving processes
• Unclear access rights based on employee need and file sensitivity	• Clear data access rights for an employee task, related to cybersecurity
• Inadequate data repurposing processes for the final intended analysis	• Upfront confirmation of data formats required for subsequent analysis step
• Incompatible analysis methodologies for the task, resulting in bias or false results	• Clearly defined outputs to business goals with the matching analysis methodologies
• Bias or incorrect understanding of the underlying data due choice of visualisation	• Accessible decision tree based on the type of visualisation required
• Failure to convert analysis into recommendations, leading to wasted insights	• Firm chain of command stipulating how insights flow towards decision makers in an organisation
• No disposal protocols or enforcement leading to potential regulatory or security breaches	• Automated disposal protocols, with regular reidentification or alignment to business needs
• Inappropriate or inadequate cybersecurity policies leading to lapses in security	• Agile and holistic security policies which are effectively communicated to employees
• Breakdown in areas of the data value chain due to weak data culture	• Tailored governance, communication, and incentive processes across the organisation
• Failure to comply with regulations resulting in regulatory fines / penalties	• Agile organisational policies which reflect and communicate changes to employees effectively

Source: Quinlan & Associates analysis

The data value chain showcases the complete lifecycle of a piece of data entering and leaving an organisation. While holistic, it certainly doesn't emphasise the gravity of the repercussions when a business gets it wrong; even if in just one part of the value chain, with the smallest misstep can have devastating ripple effects throughout an organisation.

For example, when a company fails to optimise their data input processes or create adequate policies around storage, they create what is known as "dark data." Dark data is unused, untapped, or unknown data siloed away or lost in parts of a company. This concept will be explored in-depth in our next report.

Besides financial losses, other pitfalls include reputational damage from cybersecurity breaches as a result of suboptimal security policies or processes. In addition, many companies are leaving sizeable revenue opportunities on the table as they fail to capitalise on the masses of data they are collecting; or worse, being completely unaware of how this data can support their broader business strategies. The fact is, most organisations fail to create an overarching data strategy, i.e., identify business objectives and meeting them leveraging relevant technology and data systems. And without a solid understanding of the real lingua franca of business, many companies are leaving sizeable opportunities on the table.

The next paper in this series will seek to showcase how data strategy projects flow from ideation to implementation, the stakeholders involved in each stage of the process, and its effects on a business proper.

MANY COMPANIES ARE LEAVING SIZEABLE REVENUE OPPORTUNITIES ON THE TABLE AS THEY FAIL TO CAPITALISE ON THE MASSES OF DATA THEY ARE COLLECTING; OR WORSE, BEING COMPLETELY UNAWARE OF HOW THIS DATA CAN SUPPORT THEIR BROADER BUSINESS STRATEGIES

SECTION 5
HOW CAN WE HELP?

Quinlan & Associates ("Q&A") has extensive experience working with global financial institutions, FinTechs, and other technology-based firms on their end-to-end corporate strategy development, operating model design, and implementation planning, with significant experience in the data strategy space.

Q&A's project work typically involves supporting our clients across the full data strategy spectrum:

1. DATA VALUE CHAIN AUDIT

Support companies looking to either optimise or transform their business to leverage data to pursue their business goals:

- Review existing or future business strategies and map business objectives and processes to the data value chain
- Provide a comprehensive checklist into the data value chain and how data is collected, stored, used, and disposed to meet company objectives
- Review current asset control protocols to identify areas of overspend / underspend across the data value chain
- Define the current and future state of an organisation's data systems based on its business objectives
- Perform a capability or gap analysis based on the current and future state of the business
- Review cybersecurity policies and processes in place at an organisational level and employee level
- Analyse the data culture of the organisation, including the development of comprehensive data governance and cultural policies, including employee training requirements
- Review regulatory data positioning and advise on adaptations to organisational compliance frameworks based on the company's relevant jurisdiction(s)

2. CORPORATE TRAINING

Conduct in-person or online corporate training and coaching programmes to equip our clients' employees with the necessary knowledge and capabilities to support cultural and mindset changes for a robust data value chain:

- Provide world-class employee training workshops (on areas including specific compliance topics and broader cultural change programmes), focusing on turning concepts into action, and committing actions to practice
- Engage managers and executives in dedicated coaching programmes, creating actionable plans for them to inspirate and champion good data business conduct within their teams, divisions and across the entire organisation
- Assess business performance improvements attributable to mindset and behaviour changes from training and coaching efforts, and further fine-tune the programmes.

STRATEGY WITH A DIFFERENCE

ABOUT US

Quinlan & Associates is a leading independent strategy consulting firm specialising in the financial services industry.

We are the first firm to offer end-to-end strategy consulting services. From strategy formulation to execution, to ongoing reporting, communications, and employee training, we translate cutting-edge advice into commercially executable solutions.

With our team of top-tier financial services and strategy consulting professionals and our global network of alliance partners, we give you the most up-to-date industry insights from around the world, putting you an essential step ahead of your competitors.

Quinlan & Associates. Strategy with a Difference.

enquiries@quinlanandassociates.com | www.quinlanandassociates.com | +852 2618 5000
Level 19, Two International Finance Centre, 8 Finance Street, Central, Hong Kong SAR