



Understanding the Data and Analytics Landscape in Singapore

**A Study of Data and Analytics Adoption and Practices in Six
Sectors**

Competition Commission of Singapore

KPMG Services Pte. Ltd.

16 August 2017

This report contains 84 pages



Competition Commission of Singapore

Understanding the Data and Analytics Landscape in Singapore

August 2017

Disclaimer

KPMG Services Pte Ltd (“KPMG Singapore”) has been appointed by the Competition Commission of Singapore (CCS) to conduct a study on the Data and Analytics Landscape in Singapore.

This engagement is not an assurance engagement conducted in accordance with any generally accepted assurance standards and consequently no assurance opinion is expressed. This Report contains certain information, estimates, projections, forecasts and/or assumptions (“Third Party Information”) derived from or supplied by third party sources. KPMG Singapore has not sought to establish the reliability of these sources by reference to other evidence. No representation or warranty, express or implied, is made as to the accuracy or completeness of any of this Third Party Information.

This Third Party Information is also subject to risks and uncertainties that may cause actual events and future results to be materially different from that expected or indicated in this Report and no assurance is given that any of such estimates, projections and/or forecasts will be realised. In light of these and other uncertainties, no representation or warranty is given as to the achievement or reasonableness of any estimates, projections and/or forecasts and nothing in this Report is or should be relied upon as a promise or representation as to the future.

Should this Report be relied upon by any third party for any purpose, they do so entirely at their own risk. KPMG Singapore expressly disclaims any liability (whether for negligence, under statute or otherwise) for any loss or damage of whatsoever kind (whether foreseeable or not) which may arise from any third party acting on any information or advice contained in this Report. No responsibility or liability is or will be accepted by KPMG Singapore or any of its respective directors, employees, agents or advisers in relation thereto.

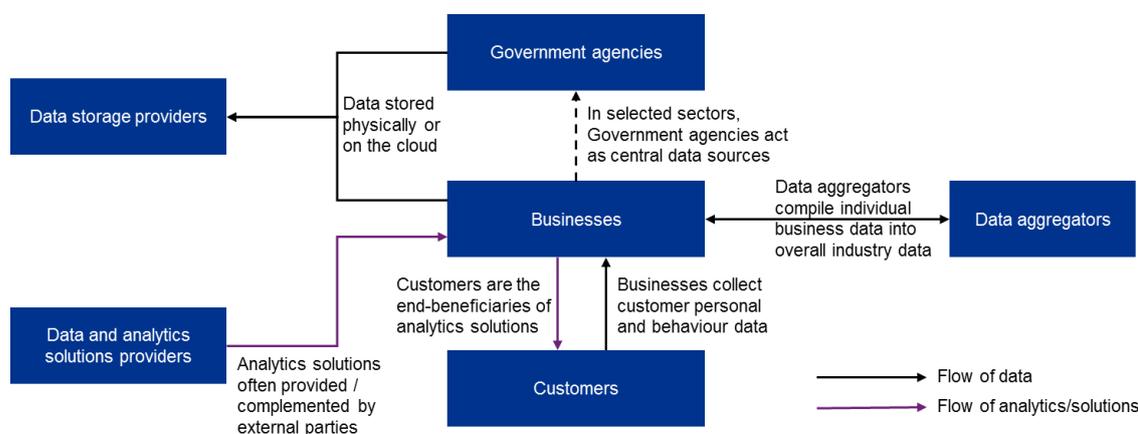
Executive Summary

KPMG has been engaged by the Competition Commission of Singapore (CCS) to undertake a study that sets out an overview of the data and analytics landscape in Singapore, through a study of six sectors. This report presents the findings from the study.

Overview of the Data and Analytics Landscape in Singapore

The main players in the data and analytics landscape are:

- **Government agencies** – which may be heavily involved in facilitating data collection and dissemination, or implementation of analytics solutions. Government involvement in data collection and analytics in a sector can significantly affect how advanced the sector is in using data analytics.
- **Businesses** – which are the backbone of the data and analytics landscape in each sector, and are the users or adopters of data analytics.
- **Customers** – who generate data for businesses, providing data on personal information, transaction habits and preferences to businesses.
- **Data analytics solutions providers** – which provide either the solutions to analytics or the tools to perform analytics. These range from companies that perform actual analytics, to software providers, mobile applications developers, and hardware developers.
- **Data storage providers** – which can be physical (within premises) or on the cloud.
- **Data aggregators** – which compile industry-level data and provide these to businesses that do not have an understanding of the overall industry.



Competition Commission of Singapore

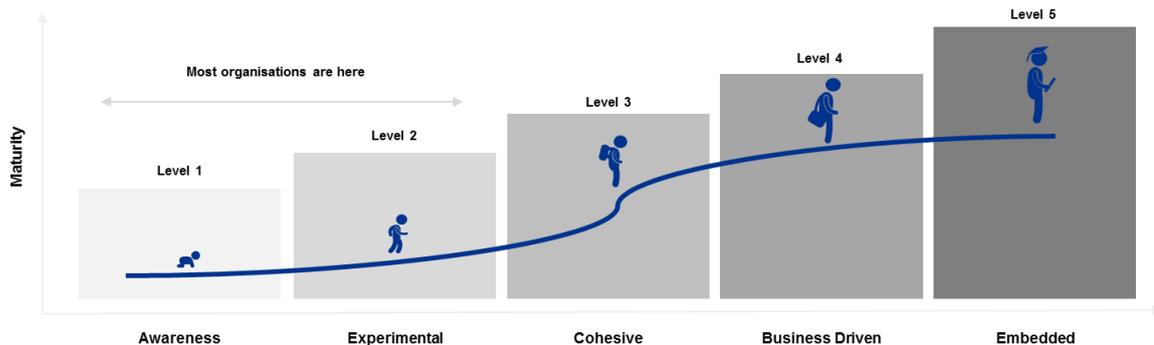
Understanding the Data and Analytics Landscape in Singapore
August 2017

There is often confusion on what analytics refers to. While there is no universally-accepted definition, analytics is largely a forward-looking and predictive mechanism while business intelligence (“BI”) focuses on historical trend analysis and patterns (e.g. what products are most commonly sold, what are customer preferences). Industry experts have made the distinction as follows:¹

- BI is needed to run the business while analytics is needed to change the business. (Pat Roche, VP of Engineering, Magnitude Software)
- BI is looking in the rear-view mirror and using historical data from one to many years ago. Analytics is looking in front of you to see what is going to happen. Both will provide you with different, not less, insights. (Mark van Rijmenam, BigData-Startups)

This distinction is important, as many businesses mistake analytics with BI. The focus of this study is on analytics.

The complexity of data analytics a company, or a sector in general, undertakes can be profiled against a maturity map. The maturity map is developed by KPMG, and comprises five stages – awareness, experimental, cohesive, business driven and embedded, from least to most mature. Most organisations and sectors are at the first two stages of maturity.



Six sectors have been identified for study: digital media, finance, healthcare, consumer retail, transport and logistics. In each of the sectors, stakeholders were interviewed to get an understanding of:

- The types of data collected and how they are used and stored.
- The types of analytics performed (e.g. predictive abilities, algorithms).
- The extent of sharing of data and analytics.
- How data and analytics are monetised (if at all).
- The impediments to greater data collection and use of analytics.
- Relevant regulations around data collection and dissemination.

¹ <https://www.betterbuys.com/bi/business-intelligence-vs-business-analytics/>

Through this, the sectors (as a whole) are assessed based on the maturity map. Findings on data analytics capabilities in each of these sectors are discussed below.

Digital Media Sector

The digital media sector is made up of online advertising platforms (e.g. Google, Facebook) and other advertising technology (ad tech) companies, which are very advanced in their use of data analytics. Customer data are widely collected; every transaction, search, product or page visited is data on customer preferences that are collected and analysed.

Ad tech companies develop technologies that help businesses organise customer data and draw insights for more effective marketing. Through all of these, analytics is used to predict customer (user) preferences, behaviour and trends.

Customer data are used in a number of ways:

- **Aggregate customers into segments** (e.g. age groups, locations) and provide advertisers with access to these specific segments (e.g. advertisers who only want to market their products to women between the ages of 20-40 in Singapore).
- **Identify and display advertisements** on search platforms that are most relevant to customers, increasing the effectiveness of the advertising campaign, based on complex algorithms.
- **Identify customers** who viewed specific products on a website but did not complete the purchase, and advertise (or “re-target”) the same products to them (e.g. websites they visit, social media accounts).

Customer data are guarded strictly by these online advertising platforms, and are not shared or sold. Data monetisation occurs indirectly through advertisement revenue; the better the quality of customer data, the more effective an advertiser’s marketing efforts, and the higher the advertising revenue generated.

Finance Sector

Finance institutions (banks and insurance) are relatively advanced in their use of data analytics. The Monetary Authority of Singapore (MAS), as the sector regulator, facilitates and encourages the use of data analytics, as part of the sector’s efforts to remain world-class.

Finance institutions have very detailed customer data, as customers are required to provide official identification and personal details (e.g. income levels, employment history, education background).

Banks are relatively more advanced in their use of data analytics, and have very developed systems that are able to collect both transactional and non-financial data on customers. Such data are used to:

- **Perform targeted marketing** to customers on products and services that they may need at different life stages and different income situations.

Competition Commission of Singapore

Understanding the Data and Analytics Landscape in Singapore

August 2017

- **Improve business operations** by identifying bottlenecks in operations, increasing talent retention and making the customer experience more efficient.
- **Automate fraud detection** to keep up with the increasing amounts of transactions undertaken.
- **Customise products** based on individual customer traits (e.g. insurance policies based on driving styles, driving distances).

Insurance companies perform similar analytics, although at a smaller scale. They are moving towards using more of the customer data collected, and incorporating official data with behavioural data (e.g. from lifestyle preferences).

FinTech and InsurTech feature prominently in the finance sector, and both MAS and financial institutions actively engage with third party collaborators to come up with solutions through technology.

Sharing of industry data is encouraged and facilitated by MAS. Customer data, especially personal identifiable information, are not shared between financial institutions. The caution around the handling of such information has sometimes led to less than optimal sharing of data, as banks are sometimes wary of sharing such data even within the organisation. Insurance companies have also cautioned against sharing customer claim histories, as it may potentially make premiums for “bad” customers so high that they become un-insurable.

Data are generally stored physically, although there is a move towards cloud storage for customer non-financial data. While MAS does not prohibit the use of cloud storage services, security concerns (e.g. cyber security) have limited the use of cloud storage.

Healthcare Sector

The healthcare sector is made up of public and private healthcare providers. Both public and private healthcare providers collect detailed patient data and medical histories, but their data analytics practices differ.

The Integrated Health Information Systems (IHIS), a wholly owned subsidiary of MOH Holdings, champions the implementation of analytics and other technologies among public healthcare providers. A prime example is the National Electronic Health Record, a central portal for patient data. The portal reduces the chances of wrong diagnoses from incomplete patient data and allows patients to avoid unnecessary duplicate tests.

Some of the analysis that public hospitals are doing are observational, although attempts at analytics and experimentation with predictive analytics is occurring:

- **Managing limited supply of inpatient capacity**, through the use of remote patient monitoring and providing healthcare services outside the hospital.
- **Optimising hospital facilities and resources** through dashboards that monitor in real-time utilisation of operating theatres and other facilities to minimise down time and understanding trends in patient demand to manage bed, doctor and nurse capacity.

Competition Commission of Singapore

Understanding the Data and Analytics Landscape in Singapore

August 2017

- **Drawing relationships** between aggregate patient health indicators with the risk of disease, and engaging patients who are deemed as high risk before they fall sick.
- **Use of predictive analytics in personalised medicine**, predicting a person's risk of contracting a critical disease in the future, based on his genetic makeup.

All public healthcare providers participate in the NEHR, but participation from private healthcare providers is more limited, although the situation is improving. Reasons for limited participation include issues with incompatible IT systems, patients' preference to keep medical conditions private and the individual practice and ownership structure of private healthcare providers.

Patient data are regarded with strict confidentiality, and only authorised healthcare professionals who are directly involved in patient care have access to the data. Patient participation in the NEHR is on an opt-out basis.

Public healthcare providers utilise cloud storage, through IHiS's Health Cloud, a private cloud platform. This allows analytics solutions to be applied more quickly and cheaply.

Consumer Retail Sector

The consumer retail sector is made up of pure e-commerce companies, as well as more traditional brick-and-mortar companies. Pure e-commerce companies are advanced in their use of analytics, incorporating both predictive analytics and other analysis of historical data. Brick-and-mortar companies are mainly only making observations from historical data, and are only becoming aware of analytics.

Pure e-commerce retailers collect a lot of data on customer behaviour and preferences, and actively use these to:

- **Improve the customer experience**, by listing popular products higher up on their websites, predict customer preferences (through favourite or similar purchases) and advertise similar products to them or by using dynamic advertising to remind customers who have viewed some products but did not complete the purchase.
- **Improve business operations**, by predicting the amount of lead time required for delivery, scaling operations through automation, managing inventory or track the performance of third party sellers (those that use their platforms).

Similar to online advertising companies, pure e-commerce retailers guard customer data strictly and treat them as their competitive advantage. Higher quality data means better customisation of products for customers and ultimately, higher sales revenue.

Pure e-commerce retailers are working towards opening up access to analytics to all employees, closing the gap for having data analytics "embedded" in its operations (the most mature stage of analytics capabilities).

Brick-and-mortar retailers generally have more limited data on their customers, as data are only collected at the transaction level, and are often not tied to individual customers. Some targeted marketing practices do occur, but this is limited by data availability.

Land Transport Sector

The land transport sector is very diverse, and ranges from private ride booking companies² (e.g. Uber, Grab) to public transport operators and private bus hiring companies. Similarly, the analytics capabilities range from highly advanced to very basic. The Land Transport Authority (LTA) drives the analytics initiatives among public bus operators, and also facilitates data sharing in the sector.

The use of analytics and respective capabilities are summarised as below:

- **Private ride booking companies** have very advanced analytics capabilities that are able to match, in real-time, demand and supply of riders.
- **LTA** utilises data visualisation techniques to map out ridership on all bus routes in Singapore, to identify commute patterns, as well as merge commuter data with video and telecommunication data for crowd control.
- **Public bus operators** have a common Fleet Management System put in place by LTA that helps in bus fleet management and scheduled adherence, as individual bus drivers can gauge the distance between it and the earlier bus, as well as compare its position to a scheduled position.
- **Private bus hire companies** are only drawing insights from historical demand and capacity utilisation data, and are only becoming aware of analytics.

LTA does not monetise the data that it shares. However, through that data, third party mobile application developers have developed applications that make real-time data available and user-friendly to public transport commuters. Private bus operators have also provided services that match commuter demand aggregated through Beeline, an online crowd-sourcing platform. These are in line with LTA's goal to make public transport easily accessible to as many people as possible.

Private ride booking companies have a lot of customer travel data that can be used to obtain travel insights. Uber is beta-testing Uber Movement, which aggregates and anonymises historical travel data to provide travel time estimates within a city.

Logistics Sector

The logistics sector has traditionally been slower in applying data analytics to its business operations, although this is expected to change with initiatives from selected leading players. Singapore's dense road network, small size and relative proximity of delivery destinations have allowed traditional business models to survive. Globally, logistics sectors are also slower in using analytics.³

Logistic service providers collect customer data when they make deliveries. However, recipient data are transient and cannot be used after the delivery is completed. Shipment data, such as those relating to carriers, international routes and costs, do not come from a central source, but are fragmented and rely on human experience.

² By definition, these companies refer to themselves as "technology companies". However, for ease of discussion, this report refers to them as ride booking companies.

³ World Economic Forum (January 2016) *Digital Transformation of Industries – Logistics Industry*

Data analytics has traditionally been limited, but this is changing. The rise of e-commerce has generated partnerships with logistics service providers to provide better delivery services and lead times. Examples of some of the uses of analytics to improve business operations, although on a limited scale by selected logistics service providers, are:

- **Warehouse automation** that makes use of autonomous vehicles and devices that have optimised routes and flow of materials.
- Collaboration with pure e-commerce companies to **predict customer demand** and deploy stock pre-emptively to a nearby location to reduce delivery lead times.
- **Historical data on delivery routes and timings** provide better estimates of delivery lead times for customers.
- **Real-time tracking** of parcels to provide delivery status updates to customers.

The Infocomm Media Development Authority has set out an Urban Logistics technology roadmap for 2020 that tests and implements new technologies to enable greater supply chain efficiencies. However, whether logistic service providers ultimately adopt the technologies is a business decision. The sector currently competes fiercely on price, and there is therefore little incentive for logistics companies to invest in analytics capabilities and infrastructure.

Impediments to Greater Use of Data Analytics and Sharing

Across sectors, stakeholders have raised the lack of skilled labour as the main impediment to greater use of data analytics. In particular, it is the lack of experienced labour with industry experience that is impeding the growth of analytics, as graduates are not provided with sufficient guidance.

Other impediments relate to the costs of setting up infrastructure and IT systems to support the use of data analytics. Individual companies do not immediately see the benefits of such investments, and often, a lot of the initiatives are Government-led.

There are also requests for there to be greater transparency and proper frameworks of the types of data that can be shared, especially those relating to personal customer details. This would facilitate greater data sharing.

Implications of Data Protection Laws

The Personal Data Protection Act in Singapore protects data that can be used to identify individuals. Before collecting, using or disclosing such data, organisations have to ensure that customer consent is obtained for that specific and reasonable purpose.

The implication of the data protection law means that businesses are very cautious in the handling of customer personal data. Such data are anonymised, encrypted and aggregated when they are used to make customer insights, and only selected employees in an organisation have access to individual customer data.

The finance sector, for example, also has additional laws around customer data protection (through the Banking Act). Strict customer data laws have meant that there is often less than optimal sharing of customer data, even within organisations. Across

Competition Commission of Singapore

Understanding the Data and Analytics Landscape in Singapore

August 2017

sectors, frameworks could be introduced to provide greater clarity on the types of data that can or cannot be shared, to improve data sharing practices.

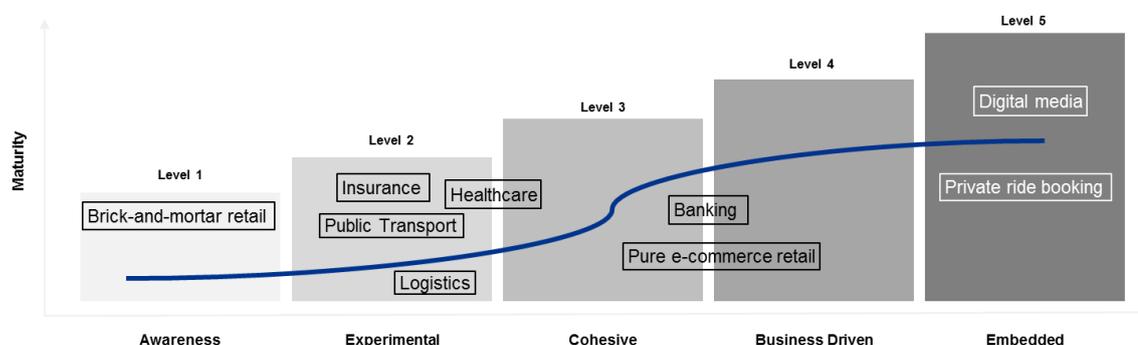
Conclusion

In conclusion, the six sectors studied have very different maturities in relation to data collection and use of analytics. The digital media sector and the private ride booking companies are the most advanced, while most sectors are only within the first two stages of the analytics maturity map.

This situation is common across organisations and similar sectors globally, and is not specific to Singapore. The global push towards collecting and using more data aims to move organisations and sectors away from the first two stages of maturity. The Singapore Government has been making significant strides in growing and enhancing the analytics capabilities in the different sectors, as described through the report.

A number of sectors, and companies within each sector, mistake analytics with BI. Making observational customer and business insights based on data is a prevalent practice (e.g. what products customers demand). As described, analytics refers to forward looking techniques with predictive abilities. True analytics, and the real value of it, are only used and realised by a handful of companies.

The maturity assessment of analytics capabilities in the six sectors is summarised in the figure below. It is important to note that these assessments are overall representations of the companies in that sector. Within any sector, there is a dispersion of analytics capabilities between companies, which in some cases can be significant.



Two key factors drive the use of and overall capabilities data analytics in sectors:

- **Government initiatives** – the efforts of Government agencies, such as LTA, MAS and IMDA, in the respective sectors have helped to ease the high costs associated with implementing and setting up analytics infrastructure and facilitated data sharing.
- **Presence of international companies** with very advanced analytics capabilities – the presence of international companies such as the private ride booking companies (e.g. Uber, Grab) and the pure e-commerce companies (e.g. Lazada) in the respective sectors in Singapore have disrupted business operations globally, and stimulated the development of analytics capabilities in those sectors in Singapore.



Competition Commission of Singapore

Understanding the Data and Analytics Landscape in Singapore

August 2017

Contents

1	Introduction	1
1.1	Background to the Study	1
1.2	Study Methodology	1
1.3	Sector Identification	2
2	Overview of the Data and Analytics Landscape in Singapore	3
2.1	Main Players in the Data and Analytics Landscape	3
2.1.1	Government Agencies	3
2.1.2	Businesses	4
2.1.3	Customers	4
2.1.4	Data Analytics Solutions Providers	4
2.1.5	Data Storage Providers	5
2.1.6	Data Aggregators	5
2.2	Personal Data Protection Act	6
2.3	Data Processes	6
2.4	Definition of Analytics	7
2.5	Data and Analytics Roadmap	7
3	Digital Media Sector	9
3.1	Data Collection and Analytics	10
3.1.1	Search Marketing	10
3.1.2	Social Marketing	11
3.2	Advertising Technology Companies	13
3.3	Data Sharing	15
3.4	Data Monetisation	15
3.5	Implications of Data Protection Laws	16
3.6	Summary of Data Analytics Maturity	16
4	Finance Sector	17
4.1	Data Collection	18
4.1.1	Collection of Personal Customer Data	18
4.1.2	Collection of Customer Behaviour Data by Banks	18
4.1.3	Collection of Customer Behaviour Data by Insurance Companies	19
4.2	Data Analytics	19
4.2.1	Using Analytics to Predict and Meet Customers' Needs	20
4.2.2	Using Analytics to Address Risk	21
4.2.3	Using Analytics to Improve Business Operations	21
4.2.4	Using Analytics to Detect Fraud	22
4.2.5	Innovations in Analytics and Technologies	23
4.3	Data Sharing	24

Competition Commission of Singapore

Understanding the Data and Analytics Landscape in Singapore

August 2017

4.3.1	Sharing of Industry Data between Banks	24
4.3.2	Sharing of Customer Data	25
4.3.3	Sharing of Customer Risk Profiles	25
4.3.4	Sharing of Product and Industry Data in the Insurance Sector	25
4.4	Data Monetisation	26
4.5	Data Storage	26
4.6	Impediments to Greater Use of Data Analytics	26
4.7	Implications of Data Protection Laws	27
4.8	Summary of Data Analytics Maturity	27
5	Healthcare Sector	28
5.1	Data Collection	29
5.2	Data Analytics	30
5.2.1	Managing Limited Supply of Inpatient Capacity	30
5.2.2	Preventive Care	32
5.2.3	Personalised Medicine	33
5.2.4	Technologies in Healthcare	33
5.3	Data Sharing	34
5.4	Data Monetisation	34
5.5	Data Storage	34
5.6	Impediments to Greater Use of Data Analytics	34
5.7	Implications of Data Protection Laws	35
5.8	Summary of Data Analytics Maturity	35
6	Consumer Retail Sector	36
6.1	Data Collection	37
6.1.1	Pure e-Commerce Companies	37
6.1.2	Other Retail Companies	37
6.2	Data Analytics	38
6.2.1	Improving Customer Experience	38
6.2.2	Improving Business Operations	40
6.2.3	Other Retail Companies	41
6.3	Data Sharing	42
6.4	Data Storage	42
6.5	Data Monetisation	42
6.6	Impediments to Greater Use of Data Analytics	42
6.7	Implications of Data Protection Laws	43
6.8	Summary of Data Analytics Maturity	43
7	Land Transport Sector	44
7.1	Data Collection	45
7.1.1	Data Collected by LTA	45
7.1.2	Data Collected by Providers of Private Ride Booking Services	45



Competition Commission of Singapore

Understanding the Data and Analytics Landscape in Singapore

August 2017

7.1.3	Data Collected by Providers of Private Bus Hire Services	45
7.1.4	Data Collected through Crowd-Sourcing	46
7.2	Data Sharing by LTA	46
7.3	Data Sharing and Collaboration between Industry Players	47
7.4	Data Analytics	47
7.4.1	Bus Fleet Management and Schedule Adherence	47
7.4.2	Public transport planning	48
7.4.3	Making Public Transport More Accessible to the General Public	49
7.4.4	Real-time Demand and Supply Matching	50
7.5	Data and Analytics Monetisation	50
7.5.1	Monetisation by Providers of Raw Data	51
7.5.2	Monetisation by Users of Data	51
7.6	Data Storage	51
7.7	Implications of Data Protection Laws	51
7.8	Summary of Data Analytics Maturity	51
8	Logistics Sector	52
8.1	Data Collection	53
8.1.1	Collection of Customer Data	53
8.1.2	Collection of Shipment Data	53
8.2	Data Analytics	54
8.2.1	IMDA's Urban Logistics Technology Roadmap	54
8.2.2	Improving Business Operations	55
8.2.3	Potential Use of Predictive Analytics to Forecast Demand	57
8.3	Data Sharing	57
8.4	Data Monetisation	57
8.5	Data Storage	57
8.6	Impediments to Greater Use of Data Analytics	57
8.7	Implications of Data Protection Laws	57
8.8	Summary of Data Analytics Maturity	58
9	Conclusion	59
9.1	Comparing Data Analytics Capabilities in the Six Sectors	59
9.2	Data Monetisation	60
9.3	Data Storage	60
9.4	Impediments to Greater Use of Data Analytics	60
9.5	Implications of Data Protection Laws	60
A	Stakeholder Survey Questionnaire	62

1 Introduction

KPMG has been engaged by the Competition Commission of Singapore (CCS) to undertake a study that sets out an overview of the data and analytics landscape in Singapore, through a study of six sectors. This report presents the findings from the study.

1.1 Background to the Study

“Big Data” and the use of it is increasingly becoming a key differentiator for many economies, as technological developments allow businesses to not only collect exponential amounts of data, but also to tap into those data and generate insights. One of the recommendations from the Committee for the Future Economy is the building of strong digital capabilities in Singapore,⁴ making use of data as an asset in business operations.

There is no single definition of Big Data, but a common understanding of it is the collection of large datasets (e.g. detailed business operations, customer behaviour and preferences) through a wide range of sources, both online and offline. The data collected may be unstructured, structured or a combination of the two, and, depending on the ability of an organisation to merge and use the data, then used to analyse and predict patterns, trends and other associations.

The rise of data as a new asset class has raised concerns on how companies use and protect the data they have, and whether the control of large amounts of data by a few companies confer some degree of market power on them. The Organisation for Economic Co-operation and Development (OECD), as well as a number of other competition agencies, have published papers on competition policy and data.⁵

1.2 Study Methodology

The study involved extensive engagement with a broad range of stakeholders between May and July 2017. This included businesses from various industries, Government agencies, third party providers of data analytics solutions, and infrastructure providers.

52 stakeholders were contacted for interview, and 27 interviews ultimately conducted, as part of the study to ensure comprehensiveness and sufficient representation of viewpoints and opinions. The survey questionnaire is included in the Appendix.

The stakeholder engagements were also complemented with KPMG’s in-house expertise and domain knowledge on the data and analytics landscape in Singapore.

⁴ See: Report of the Committee on the Future Economy – Pioneers of the Next Generation 2017

⁵ See: OECD (2016) *Big Data: Bringing Competition Policy to the Digital Era* and Bundeskartellamt, Autorite de la concurrence (May 2016) *Competition Law and Data* and JFTC’s Report of Study Group on Data and Competition Policy (June 2017)

1.3 Sector Identification

Six sectors were identified for study. The sectors were selected based on the potential of these specific sectors to reap significant benefits from the adoption of data analytics.

In each identified sector, a representative sample of companies were interviewed. The sectors, and the number of interviewed companies, are listed in Table 1.1. Companies in the “General” category refer to those that provide services that are not specific to one sector, such as providers of infrastructure, analytics solutions or data aggregators.

Table 1.1 Identified Sectors and Number of Companies Interviewed

Digital Media	Finance	Healthcare	Consumer Retail	Transport	Logistics	General
2	6	3	4	3	4	5

Within each identified sector, a mix of companies were included. The companies were of different sizes and of different maturities in their use of data analytics.

In each of the six sectors, a number of questions were asked around a number of key themes including:

- The types of data collected and how they are used and stored.
- The types of analytics performed (e.g. predictive abilities, algorithms).
- The extent of sharing of data and analytics.
- How are the data and analytics monetised (if at all).
- The impediments to greater data collection and use of analytics.
- Relevant regulations around data collection and dissemination.

2 Overview of the Data and Analytics Landscape in Singapore

This section sets out the main players in the data and analytics landscape, the types of data collected and the process through which data flows. This section also defines analytics and develops a data analytics maturity map for the identified sectors.

2.1 Main Players in the Data and Analytics Landscape

The data and analytics landscape in the different sectors in Singapore vary. However, a generalised representation is shown in Figure 2.1 below.

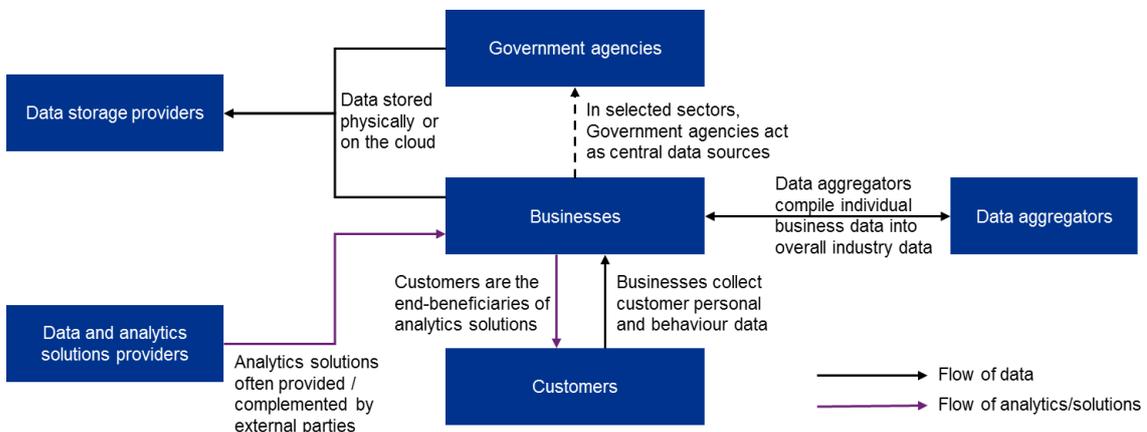


Figure 2.1 Overview of the Data and Analytics Landscape in Singapore

2.1.1 Government Agencies

Government agencies may be heavily involved in the data and analytics landscape of a sector, by making industry data publicly available or by piloting analytics solutions. Examples of such entities are:

- The Monetary Authority of Singapore (MAS) in the finance sector.
- The Land Transport Authority (LTA) in the transport sector.
- The Infocomm Media Development Authority (IMDA) in the logistics sector.

Government agencies facilitate the data sharing process by making industry data public. Examples include data.gov.sg,⁶ a central portal containing data from 70 public agencies that also makes use of data visualisation techniques to make data understandable to the public, and IMDA's initiative for a Data Marketplace, a portal with a standardised environment for public and private sector datasets.⁷ Generally, Government initiatives facilitate the sharing of non-sensitive customer data between businesses, even as they

⁶ <https://data.gov.sg/>

⁷ <https://www.imda.gov.sg/industry-development/infrastructure/next-gen-national-infocomm-infrastructure/data-marketplace>

protect sensitive data. For example, shared data can be aggregated, anonymised or encrypted to protect an individual customer's identify. Currently, there is limited sharing of data even as businesses acknowledge the mutual benefits from doing so.

Government agencies do not look to monetise data, but instead look to unlock the positive externalities of wider access to data, such as LTA's aim of enhancing commuters' public transport experience. They also try to encourage greater use of analytics in the respective sectors.

2.1.2 Businesses

Businesses are the backbone of the data and analytics landscape, and are the users or adopters of data and analytics. Businesses have access to customer data through their daily operations and interactions with customers. A business's ability to generate insights from such data and tailor their products or services to what customers want, facilitates a competitive advantage.

2.1.3 Customers

Customers contribute significantly to the data generation process, whether voluntarily or involuntarily, as it is their personal data, purchase habits, transaction histories or other behaviour patterns that allow businesses to tailor their product or service offerings.

Customers benefit from targeted marketing and customised product offering from businesses, when they are offered products that are more closely aligned with their preferences. However, it is also the ability of businesses to use these private and personal data that results in concerns around data protection and privacy.

2.1.4 Data Analytics Solutions Providers

Data analytics solution providers occupy niche positions within the data and analytics landscape. The core functions of a business is usually not in performing analytics, and as such, may require such services from external parties. These external parties provide a range of services (e.g. predicting trends, route optimisation, data visualisation), and may or may not have access to actual data (i.e. only provide software and businesses do data input, or only work with dummy data).

The data analytics industry is a key growth sector in Singapore, where demand for software solutions and professionals is expected to grow significantly.⁸ Analytics providers range from large multinational companies (e.g. Microsoft Azure, IBM Watson Analytics and SAS Analytics) to smaller companies, which often also have presence in a few cities (e.g. Aureus Analytics, Crayon Data, InfoTrie).

Third party mobile application developers also feature significantly in making data palatable and usable to end customers. These developers usually rely on publicly-available data provided free-of-charge, and therefore makes it more difficult for them to

⁸ See: <https://www.iesingapore.gov.sg/Media-Centre/News/2016/3/Opportunities-abound-in-data-analytics-sector> and <https://www.imda.gov.sg/industry-development/infrastructure/technology/data-and-analytics>

directly monetise their applications through download fees. Popular applications are, however, able to generate sizable advertising revenue.

2.1.5 Data Storage Providers

Data are stored in two main ways – physically or on remote servers accessed from the internet (i.e. cloud). Many businesses continue to store data physically, even though cloud storage is acknowledged by industry players as “the way of the future”.

Cloud storage brings benefits such as real-time data analysis (by providing a central data repository accessible by multiple users across geographies, complemented by cloud-based analytic software) and allows businesses to be nimble in their attitudes towards analytics solutions. Globally, the largest provider of public “infrastructure as a service” (or IaaS) is Amazon Web Services (AWS), through its Simple Storage Service, or S3.⁹ IBM is the largest provider of managed private cloud services.¹⁰

Cloud providers are also moving to provide cloud-based analytics services that are often pay-per-use, reducing the infrastructure investments businesses need to make. These solutions are nimble, and allow businesses to try new solutions very quickly, and move on to a new one if they are not suitable.

However, concerns have been raised with cloud storage, including the over-dependence on a single provider and the control the cloud storage provider has over the data that the business owns.

2.1.6 Data Aggregators

Data aggregators (e.g. AC Nielsen, Euromonitor) compile industry data and act as a one-stop shop for businesses that want sector-level data. The arrangements that data aggregators have with industry players vary – some data aggregators engage industry players on a quid pro quo basis, trading industry reports for insights, while others rely on individual relationships to gather insights.

Data aggregators add value to the data collection and compilation process by incorporating primary (e.g. shop surveys, interviews) and secondary research into basic company and industry data. Also, while individual businesses have data specific to themselves, data aggregators are able to extrapolate data and form insights at the industry level.

A new form of data aggregators have emerged as data brokers. Data brokers collect data on individuals (e.g. personal, demographic) through many different channels, including social media feeds, browser histories and online transactions. These data are then sold to businesses for marketing purposes, typically without customer consent. Data brokers prefer to keep their businesses and business models secretive, as they thrive on

⁹ Research by Synergy Research Group, see: <https://www.srgresearch.com/articles/amazon-dominates-public-iaas-paas-ibm-leads-managed-private-cloud>

¹⁰ Private cloud storage refers to cloud storage based from data centers that a company owns itself, as compared to public clouds that are based from data centers owned and operated by the cloud provider

customers unknowingly making data available. However, the authenticity of such data has been questioned.¹¹

2.2 Personal Data Protection Act

The Personal Data Protection Act (PDPA)¹² in Singapore protects data that can be used to identify individuals, whether accurate or not. The PDPA was set out to address concerns from individuals about how their personal data are being used as greater amounts of data are collected by organisations. There was also a need to balance the rights of individuals to protect their data and the needs of organisations to collect, use or disclose those data. The three key focus of the PDPA are:

- **Consent** – organisations are only allowed to collect, use or disclose personal data with the individual’s knowledge and consent.
- **Purpose** – organisations are only allowed to collect, use or disclose personal data if the individual has been informed of the purpose of data collection, use or disclosure.
- **Reasonableness** – organisations are only allowed to collect, use or disclose personal data for reasonable purposes.

The PDPA does not apply to Government agencies, which have their own internal protocols on data handling.

2.3 Data Processes

Data follows a three step process from collection to integration and then to storage, although the process may sometimes not be sequential, as shown in Figure 2.2 below.

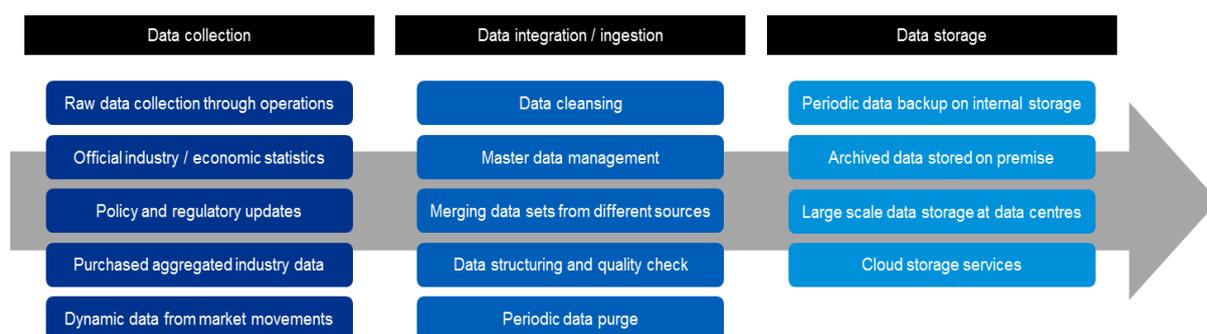


Figure 2.2 Process Flow of Data

¹¹ <http://sbr.com.sg/source/tech-in-asia/see-how-your-private-data-being-sold-without-your-permission-one-cent-apiece>

¹² <https://www.pdpc.gov.sg/legislation-and-guidelines/overview>

Analytics can be performed on data at any of these three steps, and depends on the needs of the organisation and the structure of data at that stage. The different scenarios are:

- **If analytics only requires historical data** (as compared to real-time data), then it can be performed at the data storage stage.
- **If data has to be integrated** with another set of data before insights can be drawn from them, then analytics can only be performed at or after the data integration stage.
- **If data is sufficiently structured and usable** in its raw form (i.e. the data collection and categorisation processes are complex enough), then there is no need for further ingestion before analytics can be performed.

2.4 Definition of Analytics

Analytics is a broad term, and there is often confusion on what it comprises. Analytics is different from business intelligence (BI). While there is no universally-accepted definition, BI is usually used to refer to historical trend analysis and patterns, such as in generating observational insights on current business operations (e.g. what products are most commonly sold, what are customer preferences).

In contrast, analytics is largely a forward-looking and predictive mechanism. Industry experts have made the distinction as follows:¹³

- BI is needed to run the business while analytics is needed to change the business. (Pat Roche, VP of Engineering, Magnitude Software)
- BI is looking in the rearview mirror and using historical data from one ago to many years ago. Analytics is looking in front of you to see what is going to happen. Both will provide you with different, not less, insights. (Mark van Rijmenam, BigData-Startups)

Organisations often mistake analytics for BI. BI is prevalent, while analytics less so. The focus of this study is on analytics.

2.5 Data and Analytics Roadmap

Data analytics capabilities within organisations (both businesses and Government) can be assessed against a maturity map. The maturity map shows the progression of analytics capabilities through five stages.

As shown in Figure 2.3 below, a maturity map has been developed by KPMG that sets out five stages of analytics maturities. The five stages are:

1. **Awareness** – where the organisation is only starting to become aware of data analytics. In this stage, there is no structured approach to data collection and analysis and data are usually stored in silos and may lack quality and integrity.

¹³ <https://www.betterbuys.com/bi/business-intelligence-vs-business-analytics/>

Competition Commission of Singapore

Understanding the Data and Analytics Landscape in Singapore

August 2017

2. **Experimental** – where the organisation is starting to commission and experiment with data analytics projects. In this stage, the approach to data collection and analytics is ad hoc, and used as and when the need arises.
3. **Cohesive** – where the organisation has proper data governance procedures in place and data analytics programmes are set up. In this stage, the organisation follows a structured approach to data and analytics but has yet to use insights from analytics to drive business decisions.
4. **Business driven** – where the organisation actively incorporates insights from analytics into its decision making processes. In this stage, the organisation has integrated data collection and analytics across different departments, however, enriched data is only being used by select users, such as the IT or analytics departments.
5. **Embedded** – where the organisation not only actively uses data analytics in its business decisions, but opens up access to analytics to all business areas and employees.

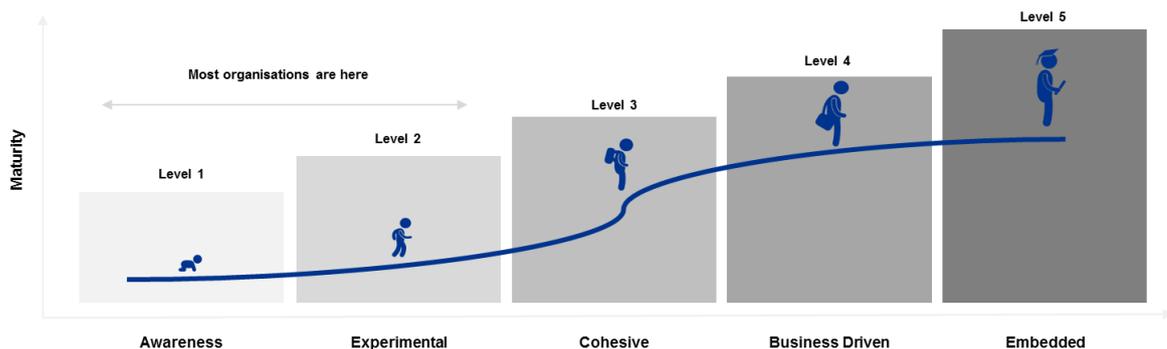


Figure 2.3 Maturity Map of Data and Analytics Capabilities within Organisations

Most sectors (and organisations within those sectors) are in the awareness and experimental stages of the maturity map. In these two stages, business insights are largely observational, and the use of analytics is ad-hoc with unstructured data collection processes.

The situation where many organisations have limited analytics capabilities is not specific to Singapore, but also generally observed globally. The global push towards collecting and using more data aims to move organisations and sectors away from the first two stages of maturity. The Singapore Government has been making significant strides in growing and enhancing the analytics capabilities in the different sectors, as described through the report.

The structure set out in this maturity map will be used to assess the six identified sectors for data analytics capabilities.

3 Digital Media Sector

The digital media sector is made up of advertising platforms and other advertising technology companies, and is very advanced in its use of data analytics. Complex algorithms and codes are designed to monitor customer behaviour, and use those data to actively predict customer preferences and intentions.

The digital media sector provides companies and organisations with a way to reach new customers, and facilitates the matching of demand (customers) and supply (businesses selling products and services) of advertisements.

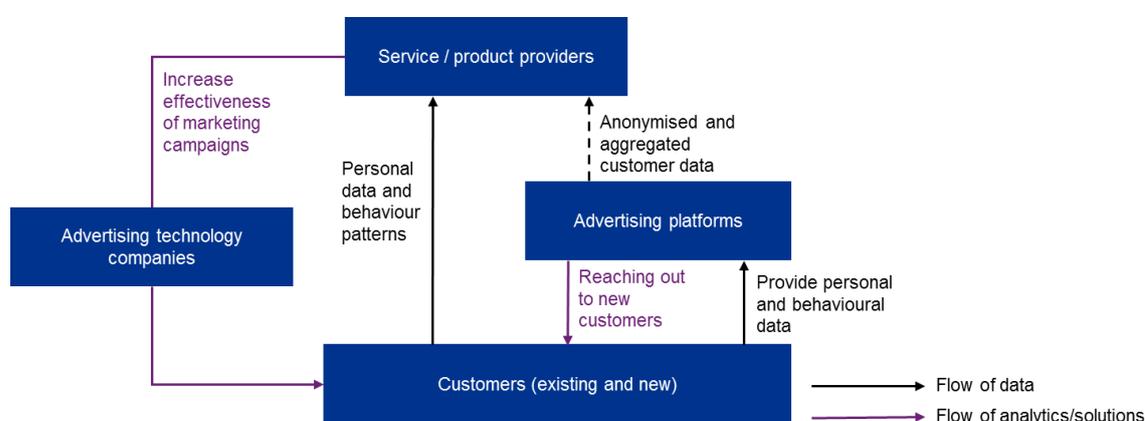


Figure 3.1 Data and Analytics Landscape in the Digital Media Sector

As shown in Figure 3.1, the main players in this sector are:

- **Customers** – who may be actively searching for products and/or services offered by businesses, or may just have passively indicated interests that are in line with what businesses are offering.
- **Service and/or product providers** – who have data on existing customers in their databases (through previous interactions) but have limited access to new customers.
- **Advertising platforms** – which facilitate the matching of these two groups of users by using collected customer data and preferences and sellers’ advertising targets.
- **Advertising technology companies (“ad tech”)** – which may or may not be part of advertising platforms, and develop technology to facilitate or improve on companies’ profiling and reach of customers.

Online advertising platforms actively collect customer data and use them to build profiles and segments of customer bases that are of value to advertisers. There are two main types of online advertising platforms:

1. **Search marketing**, through search engines (e.g. Google, Bing) when users actively search for products or services.

2. **Social marketing**, through social media applications that predict user preferences through behaviour patterns and match them to advertisers' target customers.

These main forms of marketing platforms are also known as pull (search) and push (social) marketing. The user of a search platform “pulls” information out of the platform through his search and the user of a social media platform has information “pushed” to him.

3.1 Data Collection and Analytics

The data that online advertising platforms have are generated from customer behaviour patterns on the internet (e.g. the websites they visit, the things they search for). However, more importantly, a lot of the customer data that online marketing platforms collect are provided voluntarily by customers.

As an example, users of the internet frequently “agree” to the terms and conditions of using an application or a product without fully understanding the access they are giving up to their own data. Personal data such as an individual's date of birth and gender are also easily traded for “15 minutes of free Wifi”.

Such data are valuable to the platforms collecting them, as they are used to tailor marketing efforts. The more “products” a digital media owner owns (e.g. Google also has Youtube, Gmail, Hangouts, News, Earth), the more comprehensive the data it has on an individual.

Both search and social marketing platforms widely collect and use data. Slight differences between how the two platforms work are discussed below.

3.1.1 Search Marketing

The largest search marketing platform in Singapore is Google AdWords. There are two main methods of targeted advertising (Figure 3.2):



1. **Google Search Network (GSN)**, where advertisements appear on top of / below organic search results after a user does a search query.
2. **Google Display Network (GDN)**, where advertisements appear either on the Google search page or on partner websites in the form of a banner (not shown here).

Figure 3.2 Google Search Network and Google Display Network

Source: Google AdWords

On the GSN, advertisers target customers through the use of:¹⁴

- **Keywords** – which are selected by the advertisers that are relevant to the product/service such that the advertisements appear when those terms are searched.
- **Location and language** – which can be as specific as an estate (e.g. Tampines), and searches are matched based on factors such as the user’s domain, IP address and language preference.
- **Device** – which is normally mobile or desktop.
- **Audience** – specifically, customers who had previously visited an advertiser’s site through the use of a “remarketing list for search ads”.

The same targeting techniques are used on the GDN. However, in addition, the availability of partner websites as an additional source of advertisement space allows advertisers to specify which websites the advertisements will show, through:¹⁵

- **Topic targeting** – where web content on partner pages are analysed and central themes determined (through text, language, link and page structure) to determine the suitability of display advertisement on those pages.
- **Placement targeting** – where the advertisers choose which websites to have their advertisements displayed.

Google ranks bids according to three main factors – bid amount, quality of the advertisement and the quality of user experience based on their experience at the advertised website.¹⁶ The overarching objective of AdWords is to make the most relevant advertisements available to customers, so that it closely matches what they are looking for.

3.1.2 Social Marketing

As described, social media platforms (e.g. Facebook, Twitter) collect data on their users, which are then used to customise advertisements that are fed to the users. The greater availability of user data allows advertisers to target users at a significantly higher level of granularity. In the case of Facebook, targeting can be done by:

- Location, which can be customised by people who are live in that location, who were recently in that location or who are just traveling in that location.
- Age groups.
- Gender.
- Languages used.

If businesses can complement data from Facebook with their own customer data, targeting can be even more specific, through the use of Custom Audiences or Lookalike Audiences. This is described in Figure 3.3 below.

¹⁴ <https://support.google.com/adwords/answer/1704368?hl=en>

¹⁵ <https://support.google.com/adwords/answer/1704368?hl=en>

¹⁶ Source: Google AdWords

Competition Commission of Singapore

Understanding the Data and Analytics Landscape in Singapore

August 2017

Custom Audiences		Lookalike Audiences	
Types of users targeted	Existing customers	Types of users targeted	New customers who “look like” existing ones
Data required from company	<ul style="list-style-type: none"> Traditional data from customer lists Facebook pixel, which can match visitors to a website with the Facebook user Mobile advertising identifiers, which can match users of a mobile application with the Facebook user 	Data required from company	<ul style="list-style-type: none"> Facebook pixel Mobile application data Facebook fan data
How are data used	Facebook recognises the type of data uploaded (e.g. email, phone number) and matches it to user database	How are data used	Facebook identifies common qualities of the people in the uploaded data (e.g. demographic information, interests) to find people similar to them

Figure 3.3 Overview of Facebook’s Custom Audiences and Lookalike Audiences

Source: Facebook Ads Manager

Facebook also has an advertising function called “dynamic ads”. Facebook (through its Facebook pixel) is able to:

- Identify consumers who have expressed interest on some products on a company’s website.
- Identify these consumers who then proceed to use the Facebook page / application.
- Dynamically identify the advertisements to be shown to these users.

Data requirements from an advertiser is only its product catalogue. Once an advertisement campaign is set up by the advertiser, the advertisements are automatically updated, depending on users’ behaviour and the websites visited. The advertisements are therefore “dynamic” as they change according to preferences, and “retargets” consumers.

Facebook has stated that dynamic ads are able to reach more shoppers, complete unfinished sales and contact users across mobile and desktop, regardless of where they originally shopped.

Box 1 below shows a use case of Facebook’s dynamic ads.

Box 1: ASOS's use of Facebook's Dynamic Ads to Drive Mobile Sales

ASOS is one of the world's fastest-growing online fashion retailers, offering over 75,000 product lines to customers worldwide through its country-specific websites and mobile applications. The main way ASOS wanted to grow revenue was through its rapidly expanding mobile application user base.

ASOS knew that its typical customer is a multi-platform customer who typically browses, shares and buys across a range of applications, mobile and desktop websites. Its aim was therefore to retarget advertisements to these consumers across platforms, at the right time, and with the right content.

By dynamically advertising ASOS products on Facebook to identified users, ASOS achieved the following:

35%	3X	2.5X
more people reached	increase in orders	increase in return on ad spend

Facebook also worked with ASOS to track the performance of its application, understand how consumers were interacting with products in the application and observe subsequent buying patterns. These data were used to optimise advertising campaigns

Source: Facebook Ads Manager, Success stories, ASOS

3.2 Advertising Technology Companies

The rise of the digital media sector has encouraged the emergence of advertising technology companies, otherwise known as “ad tech”. These companies develop technologies that increase the effectiveness of marketing campaigns, and have contributed significantly to the use of advanced analytics in the sector. Examples of these technologies are:

- **Data management platforms** – which are centralised platforms that combines customer data from different sources to create more accurate customer profiles that increase the effectiveness of marketing efforts. Data management platforms are useful when advertisers have advertisements across multiple networks and media owners. They are also useful for media owners to understand where demand for advertisements is coming from, as well as the top sectors. (Box 2)
- **Customer retargeting** – which serves customers advertisements based on the products or services they recently browsed online. Facebook's Dynamic Ads is an example of such technology, although other companies are also able to provide such technology.
- **Cross-device advertising** – which allows a company to advertise its products to customers across devices. With this technology, companies are able to maintain consistency in their advertisements across all media such as tablets, phones, desktops.

Competition Commission of Singapore

Understanding the Data and Analytics Landscape in Singapore
August 2017

- **Demand-side platforms** – which help marketers manage and optimise their bids for advertisement space across platforms as advertisements and advertisement space are increasingly traded through ad exchanges daily.

Box 2: Case Study on Eyeota, a Data Management Platform

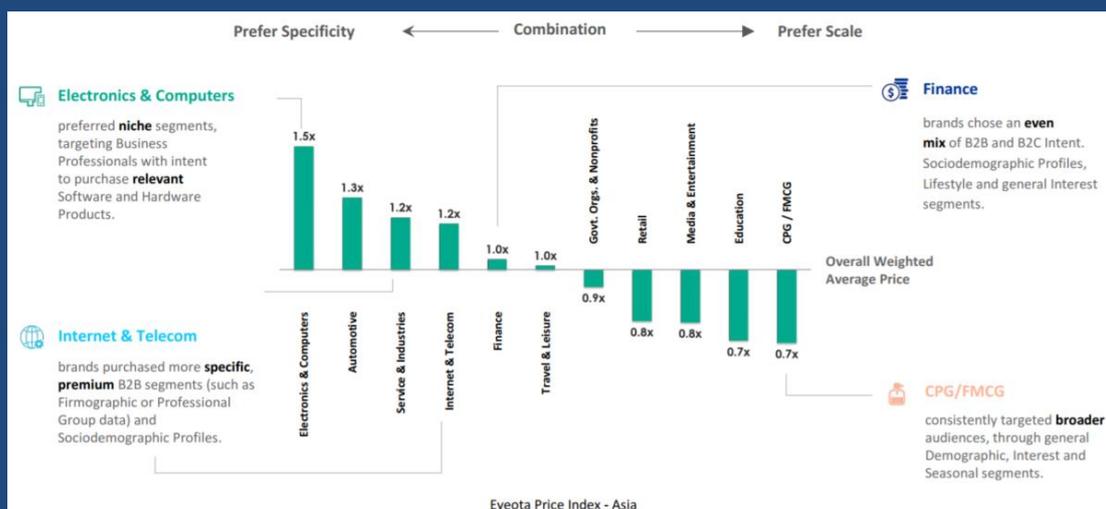
Eyeota was founded in 2010 in Singapore, and also has offices in Berlin, London, Melbourne, New York, Sydney and Tokyo. It supplies data to help advertisers more effectively segment and target consumers and expand their reach to consumers outside their network by:

- Collecting more than 3.5 billion unique profiles from over 30,000 international and local media owners.
- Pooling and aggregating these data based on socio-demographic information, interests and purchase intents on a granular level, with over 5,800 segments.
- Complementing data sets with offline data from data providers.

Eyeota has identified that Singapore represented the largest demand for data in Southeast Asia in 2016. Globally, advertising trends are that:

- Insurance brands more than doubled their audience data spend.
- Internet and telecom audience data spend grew by 2.1x year-on-year.
- Demand for millennial audience data grew by 100 per cent year-on-year.

In Asia, Eyeota found that advertisers' preferences between Scale (reach as many customers as possible) and Specificity (reach targeted customers) can be summarised in the figure below.



Source: Eyeota 2016 Annual Index Report, Your Passport to Global Audience Data Insights & Trends

3.3 Data Sharing

Customer data are never shared across platforms or media owners (e.g. Facebook, Google). Advertisers that use these platforms do not have access to customer data, except in using the data to customise its advertising audience, and even then, only at an aggregated and anonymised basis.

Platforms guard their customer data strictly, since it is the quality and exclusivity of this data that drives its value proposition to advertisers.

3.4 Data Monetisation

Advertising revenue is the key way data are monetised in the digital media sector. Advertisers bid for keywords or advertising space, and the way bidding is carried out is termed the bidding strategy.

In the case of Google AdWords, advertisers can choose between the following bidding strategies, shown in Table 3.1, which usually correspond to what their advertisements set out to do.

Table 3.1 Bidding Strategies and Advertising Campaign Objectives on Google AdWords

Bidding strategy	Description	Advertisement campaign objective
Cost-per-click	Pay for every click on the advertised search	Generate traffic to website
Cost-per-impression	Pay for every (1,000) viewable impressions, which is not necessarily a click on the advertisement	Increase brand awareness, rather than drive traffic
Cost-per-view	Pay for every view	For video advertisements to increase views or interactions with the advertisements
Conversion (AdWords Smart Bidding)	Not enough to just click or view, but a follow-on action is required	Get customers to take a direct action (e.g. sign up, download an app) on the website

Source: Google AdWords

Bidding strategies on Facebook are similar, allowing advertisers to optimise their advertising delivery by reach (advertisements will reach the maximum number of people) or impressions (advertisements will be delivered to people as many times as possible).

Competition Commission of Singapore

Understanding the Data and Analytics Landscape in Singapore
 August 2017

In addition, Facebook’s platform is able to limit advertising spend in a few other aspects, such as:¹⁷

- **Frequency cap** – which sets the number of days before an advertisement is shown to the same user again.
- **Delivery type** – which decides whether advertisements are shown throughout an advertisers’ scheduled period or accelerated (shown as quickly as possible until the advertising budget is used up).

3.5 Implications of Data Protection Laws

Customer data are never sold, but, at present, these companies are not prohibited to use existing customer data to launch new products or enter new industries.

3.6 Summary of Data Analytics Maturity

The digital media sector is very advanced in its use and adoption of data analytics. Data analytics are not only actively used to drive business decisions, they are deeply embedded in the companies’ operations. (Figure 3.4)

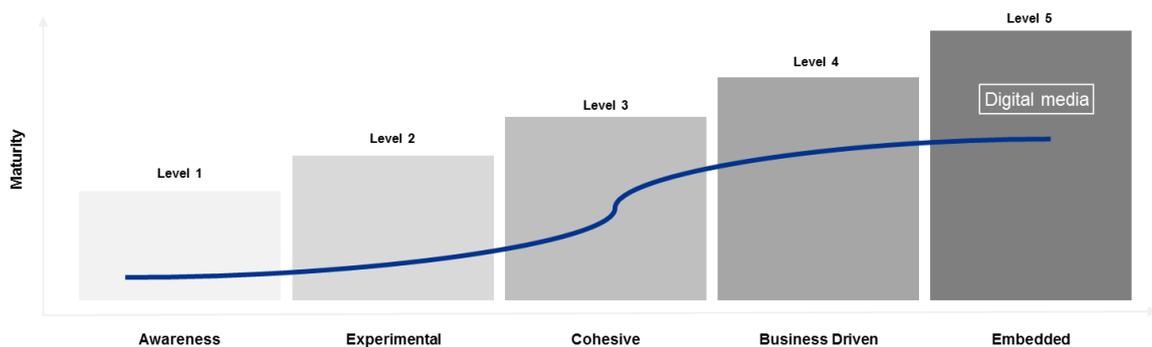


Figure 3.4 Maturity Map Assessment of the Digital Media Sector

¹⁷ Source: Facebook Ads Manager

4 Finance Sector

The finance sector refers to two broad business categories, banking and insurance.¹⁸ In Singapore, the finance sector is considered as one of the leaders of using data analytics, second only to the pure technology companies (e.g. Google, Facebook, Uber). Banks are more advanced in their use of analytics, although insurance companies also perform the same types of analytics, but at a smaller scale.

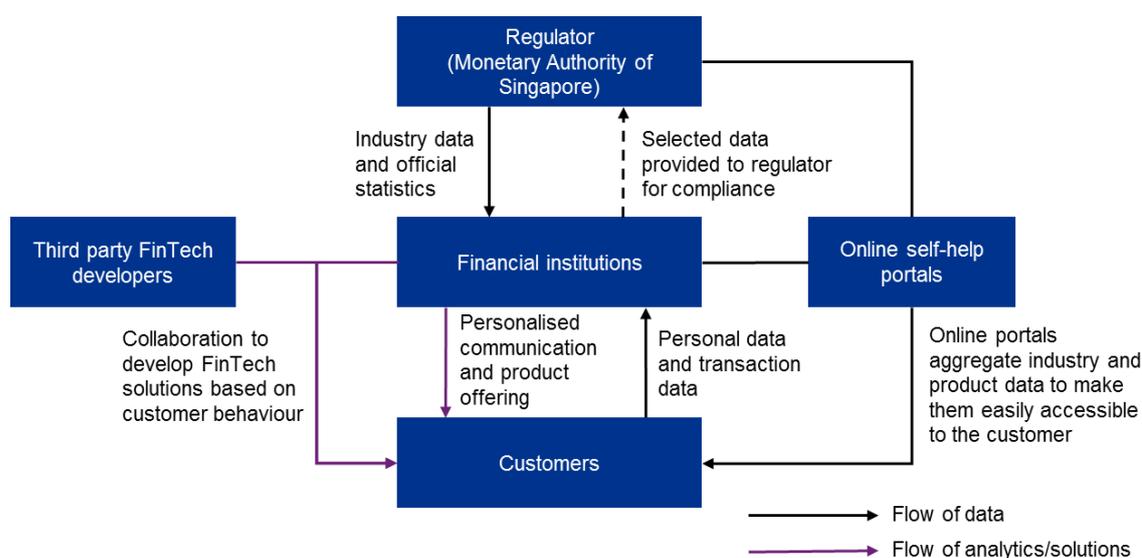


Figure 4.1 Data and Analytics Landscape in the Finance Sector

As shown in Figure 4.1, the main players in this sector are:

- **Financial institutions** – which are primarily banks and insurance companies that provide financial products.
- **Customers** – who purchase financial products and services from financial institutions.
- **Online self-help portals** – which provide customers with direct access to information on products and services that were previously accessible only through financial advisors (e.g. Fundsupermart).
- **Third party FinTech developers** – which often partner with financial institutions to provide software and other analytic solutions in the sector.
- **The Monetary Authority of Singapore (MAS)** – which is the regulator of the sector and is the source of official statistics in the sector.

¹⁸ Note: this is consistent with the definition used by MAS

4.1 Data Collection

Like all companies that are customer-facing, companies in the finance sector have access to a wealth of customer data. However, in contrast to companies in the retail sector, for example, which identify customers by only log-in email addresses, financial institutions have access to official customer identification information, as discussed below.

4.1.1 Collection of Personal Customer Data

Customers of banking and insurance companies have to provide personal details such as their NRIC/FIN, income levels, employment history, education background and sometimes, even details of family members. Such data are collected for anti-money laundering and know-your-client (KYC) processes.

In addition, banks have a fiduciary duty to collect and record data on financial transactions that customers perform. For example, information on foreign transactions, setting up an account, buying financial products, money transfers are all recorded.

Insurance companies also collect similar personal customer data. However, most of the data are often not translated into a central database. Only data that are required for ongoing management of policies are recorded; other data are left in customers' application forms, which are scanned and stored.

Personally identifiable information (PII) are only used by financial institutions when they are directly contacting customers. Access to such information is limited to selected employees within their organisation; most employees do not have access. For example, companies make use of PII when they have to:

- Send personalised emails to customers, including targeted marketing material.
- Update customers on account or product details and performance.
- Communicate company strategy or performance and direction.

4.1.2 Collection of Customer Behaviour Data by Banks

In addition to customer data, banks are also starting to collect digital data. Digital data refer to customers' footprints across a bank's network. Examples of the types of digital data collected are:

- Proportion of customers who download a bank's mobile application, frequency of use and types of activities done on the application.
- Customer preferences of branch locations, ATM location and times of contact with call centre.
- Customer behaviour on the bank's website such as the product or services they looked at, applications for products or services that were started but not completed.
- Customer behaviour and common customer activities at branches and ATMs and credit / debit card transactions.

- Customer non-financial events such as balance enquiries, seeking wealth management advice through different channels. These may be captured automatically in systems (CRM process) or manually by branch staff.

Digital data are used to complement physical data (e.g. customer activity at branches). While the amount of digital data far exceeds physical data, the real value is when the two data sets are integrated to obtain complete customer profiles.

Some banks, especially the smaller banks, purchase industry insights from external data providers to get a sense of whether its customer profile is representative of the overall industry.

4.1.3 Collection of Customer Behaviour Data by Insurance Companies

The interactions that life insurance companies traditionally have with their clients have been through insurance agents. Insurance agents are only required to collect basic customer data, and other non-standard data on customer behaviour are often not captured in official customer records. However, insurance companies are trying to change that:

- **AIA Vitality** is a wellness programme that keeps track of customers' lifestyles through its partners. AIA's customers get reward points when they attend gym or other health classes, purchase healthy food at supermarkets, quit smoking or connect their fitness devices to the programme. Customers are then put into statuses based on the points they have. AIA gets insight into customers' lifestyles through AIA Vitality.
- **NTUC Income** offers the Drive Master and FlexiMileage motor insurance schemes that allow car insurance premiums to be customised to a driver's driving behaviour and distance travelled, respectively. Through this, NTUC Income gets insights on customers' driving patterns.
- **Online and self-help insurance providers**, such as Fundsupermart (an online portal owned by iFAST) have online platforms that allow them to collect digital data, such as the transactions performed and products viewed in a more systematic and comprehensive manner.

4.2 Data Analytics

"FinTech", and the more recently coined "InsurTech", refer to the use of technologies in the finance sector. These technologies have been used to fill the gaps between what customers want and what companies offer, through an understanding of customer behaviour.

The use of data analytics in the sector is championed by MAS, which has set up its own Data Analytics Group to position itself and the sector for the digital economy of the future.¹⁹ In addition to the relatively more basic uses of analytics in improving productivity

¹⁹ <http://www.mas.gov.sg/News-and-Publications/Media-Releases/2017/MAS-Sets-up-Data-Analytics-Group.aspx>

and customer experience, analytics is also being used to detect fraud, ensure system reliability and in developing innovative products.

4.2.1 Using Analytics to Predict and Meet Customers' Needs

The push in banking companies is to personalise banks' interactions with customers. Customer feedback has been that while banks possess a wealth of information on them, there has been very limited use of that in their interactions.

Financial institutions undertake two main types of marketing techniques:

1. **Above the line marketing** – which refers to advertising on general media platforms (e.g. TV, buses/trains), and while there is some targeting (e.g. types of products, medium, roadshow events), specific audiences are difficult to target.
2. **Below the line marketing** – which refers to advertising targeted specifically to an individual, incorporating preferences and other behaviour patterns (e.g. appearing on banners when a user logs into the internet banking page, targeted statement inserts).

Generally, advertising budgets for above the line marketing are higher. However, the returns from this are difficult to track. There are numerous examples of how banks have catered their products to customers' needs, including:

- Offering credit cards with features that are catered to customer preferences (e.g. dining features and promotions as banks pick up dining-out behaviour among customers).
- Relationship managers in banks actively make use of data (e.g. transaction and enquiry histories, major life milestones) to provide a personalised experience for their customers, to make up for their inability to interact with all customers frequently.
- Analysing transaction patterns at ATMs, both to improve customer experience by allowing "favourite activities" to appear on the home page, as well as ensuring that ATMs do not run out of cash.
- Targeted marketing using customer profile and behaviour data (e.g. different investment and insurance strategies for different age groups, life stages and income levels).

Similarly, insurance companies are offering more customised services to meet their customers' needs, including:

- Starting to use sensors and wireless communication technologies in cars (i.e. telematics) to cater insurance premiums to individual driving patterns.²⁰
- Offering personalised agreements and premiums for traditionally standard insurance policies.

²⁰ <http://www.todayonline.com/singapore/car-insurers-turn-telematics-lower-risks-deter-fraud>

- Offering insurance products through their mobile platforms to cater to mobile-savvy millennial customers, streamlining the traditional underwriting process significantly (refer to Box 3 below).

Box 3: Customised and Tailored Insurance Products that Meet Customer Needs

NTUC Income provides insurance products online, targeted at customers who are used to getting access to a range of services and products with a click of a button on their mobile devices. Traditionally, insurance products are not offered online, as they have to go through a detailed underwriting process that is customised to individual customers. Technology and analytics has since allowed insurance products to be transacted online, requiring almost real-time underwriting.

NTUC Income has also introduced products that meet specific customer needs. An example of this is its Personal Mobility Guard, which addresses the gap for personal protection for alternative transport modes (e.g. bicycles, personal mobility devices).

Source: <http://www.businesstimes.com.sg/hub/fintech-2016/navigating-the-digital-landscape>

4.2.2 Using Analytics to Address Risk

Banks undertake complicated analytics to understand customer credit risks. A loan manager deciding whether to grant credit based on just the information submitted in an application form is in a poor position to decide on the credit rating of that customer. Instead, analytics can be used to provide a 360-degree view of the customer, across all touch-points and products the customer has with the bank.

In contrast, insurance companies generate customer risk profiles based on actuarial science and historical data. Premiums are created for segments of customers at the aggregate level, and not individualised. While customer behaviour data is starting to be collected, it will take some time before sufficient data are collected for it to be able to produce insights and integrated with the traditional underwriting process.

4.2.3 Using Analytics to Improve Business Operations

Analytics can improve business operations. When companies are able to map processes out from start to finish, they are able to identify bottlenecks that are slowing down the whole process. Targeted efforts on such bottlenecks can reap exponential results on productivity as compared to a more general approach.

Box 4 below shows how DBS used analytics to reduce times for calls at its call centre, reducing work load across its call centre officers.

Box 4: Interactive Analytics at Customer Centres

DBS implemented a speech analytics technology from NICE Systems that is able to analyse speech patterns on the calls DBS receives through its call centres.

Competition Commission of Singapore

Understanding the Data and Analytics Landscape in Singapore

August 2017

Calls that required longer handling times were identified and put through analytics to pinpoint recurring key phrases and determine what the causes of longer handling times were. Examples of identified reasons were unclear communications, process issues or knowledge gaps of customer service officers.

The bank found that its use of interactive analytics increased productivity, workflow capabilities and increased the number of compliments from customers by 45 per cent year on year.

Source: https://www.dbs.com/newsroom/DBS_implements_interaction_analytics_to_enhance_customer_experience_MIGRT

Talent retention is a big concern in the finance industry. Banks have partnered with analytics companies and research institutions to identify employees who have the highest risk of leaving the company in the near future. Once such employees are identified, banks can work with them pre-emptively and address any issues to lower the attrition rate.

Another area that technology is applied to, is customer authentication. The traditional customer authentication process is long and resource intensive, and technology taps into customers' unique identifiers outside of their responses to traditional questions. Box 5 below is a specific example of how OCBC uses finger and voiceprint data for customer authentication.

Box 5: Technologies used to Facilitate Operations

OCBC has incorporated several pieces of technology that are innovations on the customer process. These technologies are aimed at improving business operations, reducing the need for unnecessary customer authentication present in traditional communication methods. Examples of these are:

- OCBC OneTouch, for mobile devices, that allows customers to use their fingerprint to check their bank balances
- Voice biometrics at its call centre, where customer authentication is done through voiceprints

Source: <http://www.businesstimes.com.sg/hub/fintech-2016/navigating-the-digital-landscape>

4.2.4 Using Analytics to Detect Fraud

Fraud detection is a critical component of the operations of financial institutions. System integrity is heavily monitored and a top focus of MAS, and banks spend a lot of resources (both time and money) to ensure compliance. Technology and analytics are therefore intensely used by financial institutions to facilitate easier compliance or achieve quicker detection of non-compliance.

Box 6 below shows how analytics is used to detect anomalous transactions at DBS.

Box 6: Automated Trade Anomaly Detection

DBS partnered with A*STAR's Institute for Infocomm Research and Cloudera to develop a programme that is able to use data to detect abnormal transaction activities in the trade finance space.

Trade finance requires that banks understand its clients, the nature of the transactions, documentary financing and credit monitoring. It was a very manual and paper-intensive process as checks had to be done on every transaction.

The new system streamlined this process as it is able to identify transactional trends. Currently it works at the transaction level, but over time, there are plans to apply it at the customer level. The system has resulted in a more robust platform that is able to more automatically detect trade anomalies.

Source: <https://www.enterpriseinnovation.net/article/dbs-adopts-big-data-analytics-cut-trade-anomalies-545131553>

Fraud detection should also be inward looking, preventing conspiracies between employees to manipulate the financial system. The recent fixing of the LIBOR²¹ between employees of different financial institutions is a good example of how internal controls are equally important in ensuring a fair and reliable financial system.

Box 7 below shows an example of a software that can detect interaction patterns and establish a network of employee relationships.

Box 7: Establishing Internal Relationships to Prevent Internal Fraud

TrustSphere makes use of analytics to review internal communication data, such as the email flows within the organisation. It does not analyse the actual content of communications but the frequency of communication and response speed to establish the level of intimacy between two people. This ultimately allows a bank to map out its internal network, and detect abnormalities in relationships.

However, TrustSphere has no access to communication data, to avoid complications around data privacy. As such, frequent communications could also indicate bad relationships, and not just good ones. The software serves as a starting point rather than the end result.

Source: <http://www.bankingandfinance.com.sg/news/1165342/startup-offers-data-analytics-to-detect-internal-fraud-at-banks>

4.2.5 Innovations in Analytics and Technologies

Analytics cannot be separated from innovation and the creation of new technologies. MAS is keen to promote the FinTech landscape, but understands that concerns and uncertainty have arisen due to the nature of a regulated industry. Financial institutions may err on the side of caution if they are not clear whether a new financial product or

²¹ See: <http://www.bbc.com/news/business-19199683>

service complies with legal and regulatory requirements. This is not desirable as innovation is stifled and the sector as a whole misses out on development opportunities.

MAS has set up the FinTech Regulatory Sandbox²² which allows both financial institutions and FinTech players to experiment within a well-defined space and duration. The Sandbox also provides the appropriate safeguards to contain the consequences of failure and maintain the overall safety and soundness of the financial system. Depending on the experiment, MAS may also relax specific legal and regulatory requirements.

Big banks have also set up similar sandboxes, but within their own organisations. These are primarily due to concerns on data protection (section 4.7), and to ensure that sensitive data do not leave the organisations. Most of the big banks in Singapore have their own sandboxes, and an example of UOB's collaboration on The FinLab is shown in Box 8 below.

Box 8: UOB's The FinLab Joint Venture with FinTech Startups

The FinLab is a joint venture between UOB and SGINnovate, wholly owned by the Singapore Government under the purview of the National Research Foundation. The FinLab selects FinTech startups for its acceleration programme and provide industry experience and coaching for these new innovations.

UOB's involvement also includes assessing if the products can help serve existing clients or open up new markets in the different banking spaces. This collaboration helps to ensure that the products that are developed cater to actual needs in the industry.

Source: <http://thefinlab.com/>

4.3 Data Sharing

There are significant differences between the attitudes and practices of data sharing in relation to customer data and industry / product data. MAS as the regulator of the industry plays a significant role in encouraging and facilitating industry data sharing.

4.3.1 Sharing of Industry Data between Banks

MAS acknowledges that data analytics is key in positioning itself and the sector for the digital economy of the future. It therefore facilitates data sharing in the sector by providing real-time key data to industry players, through the use of Application Programming Interfaces (APIs). Examples of the types of data sets available through APIs are:²³

- Interest rates.
- Exchange rates.
- Asset and liability positions financial institutions.

²² <http://www.mas.gov.sg/Singapore-Financial-Centre/Smart-Financial-Centre/FinTech-Regulatory-Sandbox.aspx>

²³ <https://secure.mas.gov.sg/api/Search.aspx>

- Loans and advances by financial institutions.

Banks also develop their own APIs for external parties to access, such as OCBC's developer portal (Connect2OCBC). The portal houses a number of APIs that developers can request for access to, so that information on OCBC's products and services can be easily integrated into third party applications to benefit end-users.²⁴

4.3.2 Sharing of Customer Data

Customer data are not shared between companies. This is due to regulations around data privacy and particularly for banks, the secrecy provisions in the Banking Act.²⁵ Banks are not allowed to disclose the identity of their customers, and this includes the matching of customer identifiers such as NRIC numbers with the databases of other companies, even if the identifiers are encrypted.

Larger companies have less incentive to share data as the data they already own can be more easily extrapolated to generate industry insights.

4.3.3 Sharing of Customer Risk Profiles

Customer risk profiles and histories are not shared between insurance companies. If such data are shared, customers will be affected in two ways:

- Customers with a good history of few medical claims will be rewarded and face lower premiums.
- Customers with a bad history will be penalised, and face very high premiums or not even be insured by any company.

Overall, the net benefit to customers from sharing of such data is ambiguous although insurance companies have raised the concern of making some customer un-insurable.

4.3.4 Sharing of Product and Industry Data in the Insurance Sector

Traditionally, there has been minimal sharing of product and industry data between insurance providers and customers. The only way a customer could get information on insurance products was through an insurance agent or financial advisor, and customer relationships are fiercely protected.

However, this situation is changing. Insurance companies are beginning to go online, and insurers such as NTUC Income, for example, have a mobile platform where customers can get direct quotes for insurance products without going through an agent. Companies such as Direct Asia are also cutting out the middleman and facilitating direct and transparent interactions with customers.

Companies such as iFAST are bridging the information gap between product providers and customers, both for industry and product data. Its business model relies on the

²⁴ <https://api.ocbc.com/store/>

²⁵ See: Section 47 on Banking Secrecy of the Banking Act 2008 that says "*Customer information shall not, in any way, be disclosed by a bank in Singapore or any of its officers to any other person except as expressly provided in this Act.*"

internet to change the way insurance products are sold by making research and information widely available to the general public on its website.

4.4 Data Monetisation

Both customer and industry data are not monetised directly (i.e. sold), but through indirect means such as:

- The use of targeted advertising ensures that only the products and services that a customer needs are pushed to them, increasing the effectiveness of the marketing campaign and conversion rates.
- The use of personalised communication increases customer loyalty.
- The opening of product information (e.g. through APIs) to third party application developers helps to increase the marketing reach of a financial institution.

4.5 Data Storage

The data collected by financial institutions are generally stored on premise, although there is a move towards cloud storage for customer non-financial data. While MAS does not prohibit the use of cloud storage services, security concerns (e.g. cyber security) have limited the use of cloud storage. Current data storage systems are set up to comply with provisions around data security, for example, and moving to a new platform means re-doing the process. There are also concerns of over-reliance on a single cloud provider, and the cloud provider having control over a company's data.

Businesses' data storage policies are also affected by the regulations in the different countries they operate in. If the regulations in other countries do not allow cloud storage services, businesses tend to prefer a standard data storage policy across all its branches.

An example of such a regulation is the localisation rule. In Indonesia, Regulation No. 82 of 2012 requires, among other things, that companies with operations in Indonesia have data centres and IT systems that are hosted in Indonesia.²⁶ Banks with Indonesian operations will therefore have to comply with this regulation, and it may therefore not make sense for a partial data storage strategy that applies to a subset of the countries the banks operate in.

4.6 Impediments to Greater Use of Data Analytics

The main impediment to greater use of data analytics is the lack of human resource. Specifically, it is the lack of experienced manpower in analytics (with about 10-12 years of experience) that is limiting the sector. Businesses are adjusting to this by importing talent, as they wait for Singapore's supply of talent to mature in the future.

Insurance companies are also in the process of adapting their data collection process to record and monitor more customer data. As described above, a lot of customer data remains as hard-coded data that are not recorded in a central database. Insurance

²⁶ http://www.flevin.com/id/lgso/translations/JICA%20Mirror/english/4902_PP_82_2012_e.html

companies are also in the process of integrating customer behavioural data with traditional historical data used for underwriting.

4.7 Implications of Data Protection Laws

Regulations around data protection are stringent in the finance sector because, in addition to the Personal Data Protection Act (PDPA), there are industry specific rules (e.g. Banking Act). Financial institutions are cautious of their handling of customer data, especially those on PII. Often, such rules have limited the sharing of data, even internally, as these institutions prefer to err on the side of caution and take the strict interpretation of the law.

An industry wide framework on data classification would reduce the uncertainties around this area, which sets out the types of data that can or cannot be shared. This will greatly facilitate greater sharing of data between financial institutions.

4.8 Summary of Data Analytics Maturity

The finance sector is relatively advanced in its use of data analytics, although there are some differences in the level of maturity between companies. Banks have cohesive frameworks for data collection, and are using data to drive business decision. Insurance companies, on the other hand, are only starting to experiment with data analytics. (Figure 4.2)

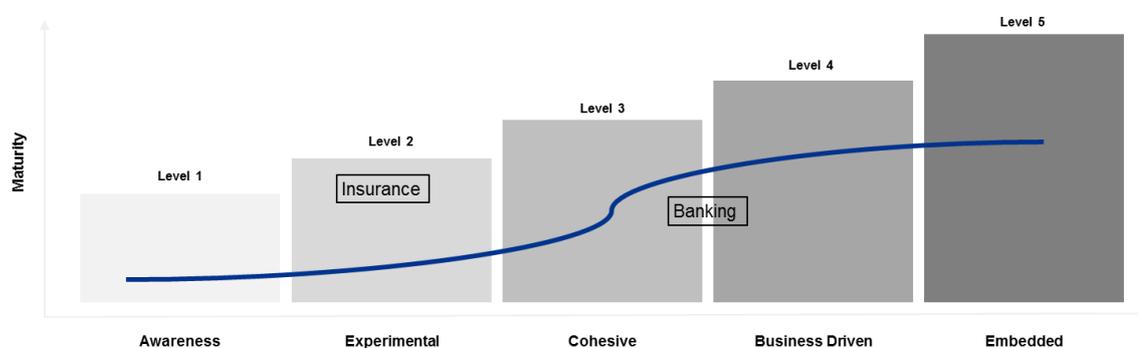


Figure 4.2 Maturity Map Assessment of the Finance Sector

5 Healthcare Sector

The healthcare sector in Singapore is burdened by an ageing population and an increased chronic disease burden. Lowering healthcare costs and managing limited capacity are two main areas of focus in the sector.

Efforts on facilitating data collection and sharing by the Government have been implemented in public healthcare providers, but to a smaller extent in private healthcare providers. Some of the current analysis of data is based on making observations of historical data, although there are attempts with using and applying analytics.

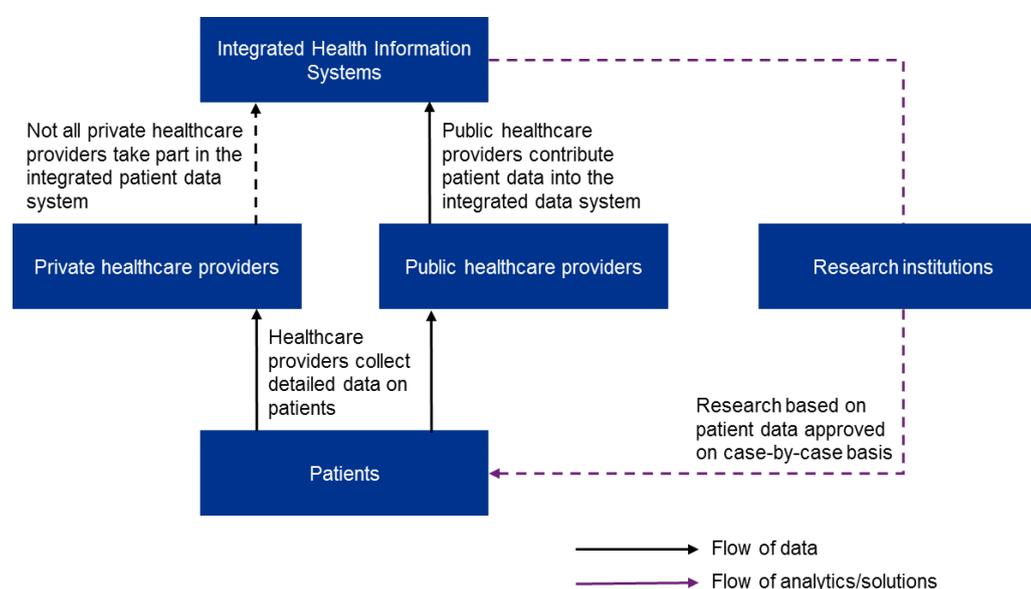


Figure 5.1 Data and Analytics Landscape in the Healthcare Sector

As shown in Figure 5.1, the main players in the healthcare sector are:

- **The Integrated Health Information Systems (IHIS)** – which manages the integrated systems and IT expertise across Singapore’s public healthcare providers and operates the National Electronic Health Records (NEHR), a central portal that stores patient data. It is wholly owned by MOH Holdings, a holding company that owns all the public hospitals in Singapore.
- **Public healthcare providers** – which make up about 20 per cent of primary healthcare providers (polyclinics), 80 per cent of secondary and tertiary health providers (public regional hospitals and specialty centres) and 70 per cent of intermediate and long term care providers (voluntary community hospitals, hospices and nursing homes).²⁷
- **Private healthcare providers** – which make up about 80 per cent of primary healthcare providers (GP clinics), 20 per cent of secondary and tertiary health

²⁷ MOH Holdings *Singapore’s National Electronic Health Record (NEHR) – The Journey to 2012 and Beyond*

providers (private hospitals) and 30 per cent of intermediate and long term care providers.

- **Research institutions** – which are typically part of a public healthcare cluster and undertake health research and analytics work for all public hospitals within the cluster. They have limited access to patient data, and have to comply with internal regulations on the use of data.
- **Patients** – who use the healthcare system to receive primary, secondary, tertiary and term care healthcare services and provide medical histories as data.

5.1 Data Collection

There is detailed collection of patient data every time a patient visits any healthcare provider. Clinics and hospitals require official identification before they will see a patient, and other types of documentation are also provided, such as subsidy cards and medical insurance coverage cards.

When doctors attend to patients, they take detailed clinical notes, recorded in a patient's medical record. Such records are electronic, and are referred to as patients' Electronic Medical Records (EMR). EMRs record the process a doctor takes in providing a diagnosis of a patient, and are used by both private and public healthcare providers.

The National Electronic Health Records (NEHR) is a centralised patient data system operated by IHiS and has the vision of "One Patient, One Health Record" for Singapore. It aims to facilitate the seamless flow of information and help integrate medical care across different healthcare settings in Singapore. The NEHR records patients' medical histories, including:

- | | |
|--|--|
| ▪ Admission and visit history | ▪ Medication history |
| ▪ Laboratory results | ▪ History of past observations |
| ▪ Radiology results | ▪ Immunisations |
| ▪ Hospital inpatient discharge summaries | ▪ Allergies and adverse drug reactions |

Unlike the EMR, the NEHR only records the final diagnosis of a patient. The benefits of the NEHR are experienced by both healthcare providers and patients:

- The NEHR reduces the chances of inaccurate diagnoses. Without the NEHR, doctors only relied on patients' responses to get an indication of the medication they are on and their past diagnoses. Patients may forget what they have been given, or just do not completely understand their conditions, and therefore give an incomplete summary to the doctor, leading to sub-optimal treatment.
- The NEHR helps to reduce healthcare costs for patients by avoiding unnecessary or duplicate tests when patients switched healthcare providers.

Competition Commission of Singapore

Understanding the Data and Analytics Landscape in Singapore

August 2017

All public healthcare providers participate in the NEHR, but participation from private healthcare providers is more limited, although the situation is improving.²⁸ A variety of reasons are associated with this limited participation, including:

- Issues with compatibility and capabilities of IT systems and the impact on operational costs.²⁹
- Patients' preference to keep their medical conditions private (e.g. sexually transmitted diseases).³⁰
- Individually-owned private practices aggregated under a common brand without a unified approach to investment in IT systems.

Patient participation in the NEHR is on an opt-out basis. Patients who opt out of the system will still have their data collected, just not uploaded onto the system, so that no additional data have to be collected if he decides to opt in in the future.

Individual hospitals are custodians of patient data. Only professionals who are directly involved in patient care have access to that data; even research departments within the hospitals, for example, do not have direct access.

5.2 Data Analytics

IHiS has set up a coherent framework for collecting data, although the use of analytics in the sector is between the experimental and cohesive stages. The use of analytics is more prevalent among public healthcare providers, due to initiatives from IHiS.

5.2.1 Managing Limited Supply of Inpatient Capacity

The growing burden of chronic diseases means that the supply of inpatient capacity has to be carefully managed, rather than simply expanded as that would lead to increased costs on patients. Historically, hospitals have found that not all patients who visit the hospitals' Accident & Emergency (A&E) departments or request to be admitted necessarily require that. If hospitals can identify and reduce the number of patients who do not necessarily need to be there, they can treat the patients who do, more quickly and effectively.

Some of the ways this is being done include:

- Monitoring patients remotely to identify a deterioration in health before it becomes critical, or deploying transitional care teams that provide non-emergency healthcare services to patient at their homes. (Box 9)

²⁸ As of April 2017, more than 21,000 healthcare professionals, including doctors, nurses and pharmacists, from more than 1,000 healthcare providers have access to the NEHR. In a typical month, more than 767,000 patient record searches are being made. For a list of institutions participating in the NEHR, see: https://www.moh.gov.sg/content/moh_web/home/national-electronic-health-records-listing.html

²⁹ https://www.moh.gov.sg/content/moh_web/home/pressRoom/Parliamentary_QA/2015/gps-utilisation-of-nehr.html

³⁰ Straits Times article on 3 July 2017 "MOH may make it mandatory for private sector to join national health database"

Competition Commission of Singapore

Understanding the Data and Analytics Landscape in Singapore

August 2017

- Having oversight of facilities to minimise down time and inefficient scheduling such that the same facility can treat more patients in the same amount of time. (Box 10)
- Bed occupancy data used to identify trends in increased demand patterns, to allocate beds optimally across different departments or provide an estimate to patients on waiting time for beds. This is to avoid the situation of a bed crunch.
- Tweaking the rosters of doctors and other healthcare providers to more closely match periods with increased patient demand. (Box 11)

Box 9: Remote patient monitoring and care

Changi General Hospital (CGH) set up a tele-health programme for heart failure patients in Singapore, who have been found to require frequent re-admission. In 2014, the hospital had a 40 per cent re-admission rate within 12 months. A key reason for the high re-admissions rate was non-adherence to prescribed treatment plans (e.g. not taking medicine correctly / consistently, not monitoring weight daily, lapses in control of water intake and diet).

The programme was put in place through a joint effort between the Eastern Health Alliance, CGG and Philips Healthcare. The programme integrates three key elements:

1. **Tele-monitoring** – where patients were provided with equipment to monitor their health (e.g. weight, blood pressure) with statistics automatically uploaded to a central system for monitoring.
2. **Tele-education** – where educational materials such as videos are delivered remotely through patients' issued tablets.
3. **Tele-care** – where nurse telecarers provide support and monitoring, and intervene early when signs of deterioration are detected.

In another example, Khoo Teck Puat Hospital (KTPH) found that out of 21,000 patients, 425 were admitted three or more times over six months, taking up 10,000 bed days (equivalent to 11 per cent of total available bed days). The hospital formed an Ageing-in-Place team, which included nurses and part-time social workers to help patients receive care at home. KTPH had a 47 per cent reduction of the chronically ill being readmitted, freeing 5,000 bed days in a six month period.

Source: <http://www.easternhealth.sg/Pages/MRDDetails.aspx?ProjectID=16> and <http://sbr.com.sg/healthcare/commentary/data-analytics-alexandra-health-system-new-journey-in-healthcare>

Box 10: Real-time monitoring of facility use and capacity

In 2013, KTPH faced challenges in meeting the demand for A&E Care Centre services, exceeding its intended patient capacity by 50 patients a day. A simple solution may have been to just hire more staff, but that was a short term solution that puts more strain on managing costs.

Competition Commission of Singapore

Understanding the Data and Analytics Landscape in Singapore

August 2017

To cope with this, the hospital implemented operating theatre dashboards that monitored its 14 operating theatres for cancellations, overruns and late starts. The dashboards also adopted key performance indicators from US hospitals, such as the utilisation rate of operating theatres and their daily schedule to provide an overview of available resources. These were monitored regularly to enhance the hospital's process efficiency.

Source: <http://sbr.com.sg/healthcare/commentary/data-analytics-alexandra-health-system-new-journey-in-healthcare>

Box 11: Adjusting Roster Schedules to Coincide with Patient Demand

CGH analysed patient data in an attempt to manage resources. It proposed a redistribution of manpower, arranging for doctors' shifts to have greater overlaps during the peak hours of around 10 am and 7 pm to 8 pm. The benefits were:

- 24 per cent improvement (33 minutes to 25 minutes) in average median time to first consultation for patients with more serious conditions.
- Doctors reported a qualitative improvement in workload.

Source:

https://www.ihis.com.sg/Latest_News/News_Article/Pages/Big_Data_Making_A_Great_Difference_in_Healthcare_TODAY.aspx

Related to remote patient monitoring, the Internet of Things will feature prominently in future healthcare scenes in Singapore. The *Future of Us* exhibition that ran from December 2016 to March 2016 is a display of what life could be like in 2030.³¹ In the healthcare sector, it featured the use of smart watches and wristbands that can detect an elderly person's vital signs. The devices are linked to a healthcare system, allowing community healthcare volunteers or doctors to be alerted, who can then reach out to the elderly person quickly.

5.2.2 Preventive Care

The most ideal scenario for patients is prevention. If patients could be warned before they fell sick, they could take steps to remedy the situation, and maybe avoid falling sick altogether.

Hospitals are experimenting with analytical models that predict a patient's likelihood of disease (e.g. chance of renal failure). These models are built based on aggregate patient data and health indicators, and are mostly backward looking:

- Historical relationships between a patient's lifestyle and the chance of disease are identified.
- Indicators that tend to increase the chance of disease within the Singapore population are identified.

³¹ Source: <https://www.gov.sg/microsites/future-economy/press-room/news/content/say-hello-to-the-internet-of-things>

People who possess the high-risk indicators are actively engaged with to try and mitigate the risk of disease before it starts. Box 12 below is an example of such a practice.

Box 12: Preventive Care Modelling Exercise

The Alexandra Health System (AHS) has piloted the Population Health Programme, which attempts to identify illness before they happen. To do this, AHS collected data demographic and clinical data to stratify the population into various health groups. Using a basket of indicators, it identified the at-risk residents and worked to improve that through health education and health promoting activities.

Source: <https://www.imda.gov.sg/infocomm-and-media-news/buzz-central/2014/7/healthcare-analytics-in-action>

5.2.3 Personalised Medicine

The sector is also experimenting with personalised medicine, where a patient's genetic makeup can be used to predict the likelihood of a serious disease in the future, possible risks of complications from treatments or the efficacy of antibiotics, for example. This area of analytics is purely research-based, and is still at an infancy stage in Singapore.

5.2.4 Technologies in Healthcare

IHiS is pioneering the implementation of advanced and other technologies in the healthcare sector. It currently manages 780 IT and other systems that support more than 50,000 healthcare professionals, and has won 80 awards for its initiatives. While these technologies currently apply primarily to public healthcare providers, they will naturally spread through the entire sector over time.

Examples of the systems that IHiS has put in place are:³²

- An **integrated queue and payment system** (1Q1P) that generates on slip with one queue number for a patient's entire set of appointments per visit, and a consolidated payment at the end.
- A **data consolidation and dashboard reporting system** that collects, cleanses and standardises data across functions of a hospital's day-to-day operations.
- A **medical device integration solution** that automatically collects and integrates patients' vital signs data generated from various medical devices to the NEHR wirelessly.
- A **multi-dose medication management system** that automatically dispenses patients' medication into a series of single sachets, saving nurses' time in re-packaging medication.
- **Mobile phone applications** for asthma patients (real-time instructions on what to do and medication dosages), anxiety patients (manage panic attacks and set goals for recovery) or diabetes patients (insulin dosage based on blood sugar reading).

³² https://www.ihis.com.sg/Project_Showcase/Healthcare_Systems

5.3 Data Sharing

Access to patient data is guarded very strictly. Only doctors and healthcare providers who are directly involved in a patient's care have access to that patient's data. Doctors also have access to individual patient data when their opinions are sought as a second consult.

Patient data are not shared with stakeholders who are not involved in patient care. However, there have been instances of unauthorised access of patient data that resulted in the doctor or healthcare professional being punished for it.³³

Data that are used for research are approved on a case-by-case basis, and have to comply with internal procedures, requirements as well as ethics guidelines.

5.4 Data Monetisation

IHiS and public healthcare providers do not seek to monetise the analytics solutions they possess. There are greater aims like improving the quality of patient care, and ensuring that the healthcare system is seamless and integrated.

5.5 Data Storage

Patients' electronic medical records are stored on a private cloud platform, termed Health Cloud or H-Cloud.³⁴ H-Cloud is developed by IHiS as an agile yet cost-effective cloud environment that would keep the healthcare infrastructure robust and enable healthcare institutions to deploy application systems more cheaply and quickly. It won the *DataCloud Enterprise Cloud Award: Cloud End User Innovation Award* in 2015.

Through the use of H-Cloud, infrastructure and application availability was increased from 99.5 per cent to 99.99 per cent. Public hospitals would also realise significant savings from using H-Cloud, by about 55 per cent by 2025.

There are plans to introduce the use of cloud storage to private healthcare providers. The Infocomm Media Development Authority mapped out cloud security standards to private healthcare providers in 2016, to facilitate their movement to cloud storage.³⁵

5.6 Impediments to Greater Use of Data Analytics

As with other sectors, the largest impediment to greater use of analytics in the healthcare sector is the lack of skilled labour. The healthcare sector is often not a top choice of employment for experienced data scientists.

While the Government has started many initiatives for more data sharing and use of analytics in the sector, private healthcare providers may sometimes be unwilling to invest in analytics solutions as the returns are not immediately felt.

³³ <https://www.sma.org.sg/UploadedImg/files/Publications%20-%20SMA%20News/4906/CN%203.pdf>

³⁴ See: IHiS media release (6 June 2015) *Singapore's Health Cloud Edged Out Hundreds of Global Submissions to Snag Prestigious DataCloud Enterprise Cloud Award in Monaco*

³⁵ <https://www.imda.gov.sg/about/newsroom/archived/ida/media-releases/2016/singapore-maps-cloud-security-standards-to-private-healthcare-sector-exploring-data-certification-framework>

Unlike public healthcare providers, which are focused on making healthcare cost-effective for patients, private healthcare providers are generally less concerned with costs due to different patient profiles. Instead, private healthcare providers focus on providing better levels of service for their patients, which may not rely as heavily on data analytics.

5.7 Implications of Data Protection Laws

The NEHR contains patient treatment data, but not other customer data (e.g. personal data, Medisave balance and applicable assistance and subsidy schemes).

Sometimes, healthcare providers may want access to these other customer data, primarily to determine whether Government subsidies/assistance or the amount of Medisave and Medisave Life coverage that are applicable. In these instances, patients have to provide consent for such data to be shared. When consent is given, patient data can only be accessed for official purposes, on a need-to-know basis by the Government and other approved organisations.³⁶

5.8 Summary of Data Analytics Maturity

Overall, the data analytics capabilities in the healthcare sector are assessed as in between the experimental and cohesive stages. While IHiS has set up a proper structure for collecting and sharing patient data, not all healthcare providers are on board with the initiative yet. Most of the analysis undertaken are backward looking, and predictive analytics (e.g. personalised medicine) is still at an infancy stage. (Figure 5.2)

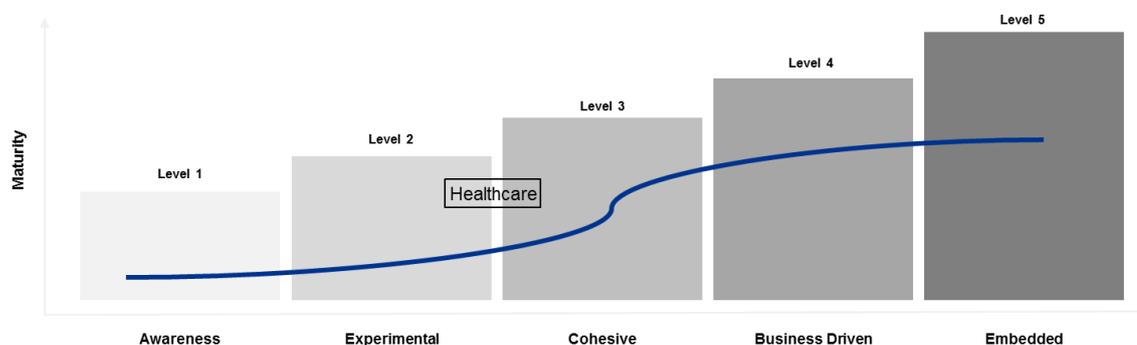


Figure 5.2 Maturity Map Assessment of the Healthcare Sector

³⁶ https://www.moh.gov.sg/content/moh_web/home/costs_and_financing/data-sharing.html#1

6 Consumer Retail Sector

The consumer retail sector is broadly divided into two main segments – online retail and physical (brick-and-mortar) retail. The lines between the two segments have blurred in recent years, with physical retail having some online presence, and (smaller) e-commerce platforms expanding into the physical space.

As examples, traditionally physical stores such as Charles and Keith and Popular Bookstore are starting to have online business units, whereas online stores such as blogshops (e.g. MDS) are starting to have physical stores.

Pure e-commerce companies are the undisputed leaders in their use of data analytics in this sector.

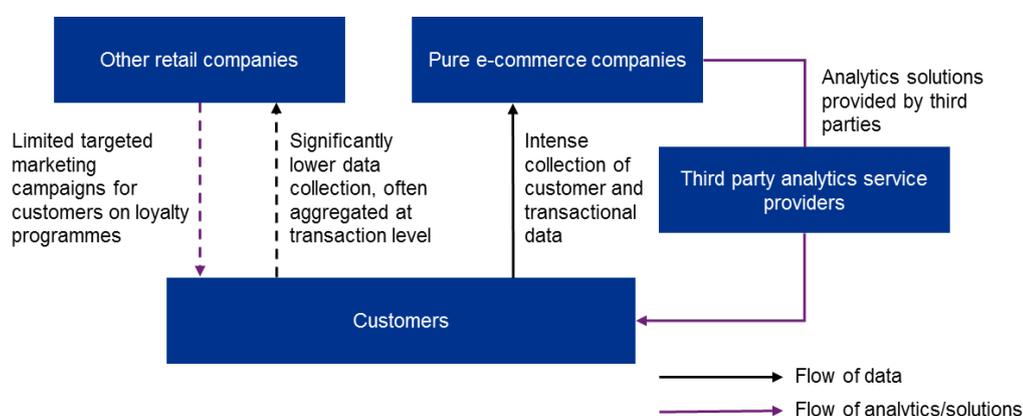


Figure 6.1 Data and Analytics Landscape in the Consumer Retail Sector

As shown in Figure 6.1, the main players in this sector are:

- **Customers** – who are the shoppers and purchase products offered by retailers.
- **Pure e-commerce companies** – which conduct all of their business online and intensely collect and use data, such as Lazada, Amazon and Zalora.
- **Other retail companies** – which may have either or both of an online and physical presence.
- **Third party analytics service providers** – which provide analytics solutions especially to pure e-commerce companies, and often do not have access to any company data.

The pure e-commerce companies are typically two-sided platforms, bringing together two groups of stakeholders:

- **Customer front** – which refers to the users of the platforms who use the platforms to purchase goods and services.
- **Seller front** – which, depending on the companies' business models, can refer to the third party sellers that use the platforms to sell goods and services to

customers and / or internal operations that perform the same functions to customers.

6.1 Data Collection

In all retail companies, data are collected around customers and operations. However, the amount of data collected varies.

6.1.1 Pure e-Commerce Companies

In pure e-commerce companies, vast amounts of data are collected from daily operations and customer transactions, and the amount is continuing to grow over time. The nature of their operations mean that two main types of data are collected:

1. **Customer data** – which pertains to customer activity (e.g. who logs into their account, where that is), customer behaviour (what pages and products customers visited and viewed, how long customers stay on a page), transaction history (including transactions that drop out before the final checkout and payment).
2. **Merchant data** – which pertains to merchant responsiveness, merchants' ability to fulfil orders, reviews from customers on merchants.

These data are generated internally through day-to-day operations, with generally limited purchase of customer or merchant data from external sources. The nature and the specificity of data to a company's operations necessitate internal data generation.

Typically, the same types of data are collected across platforms (e.g. mobile, desktop) and across sellers (e.g. merchants, own retail).

Other than data relating to customers, data that are general to the industry (e.g. market trends, shares, growth rates) are obtained from data aggregators. Research companies such as AC Nielsen and Bloomberg provide these data, and arrangements can vary, as described in Section 2.1.6.

6.1.2 Other Retail Companies

Other retail companies also collect data on customer behaviour and transactions. However, the scope and breadth of the data collected is much more limited.

These retail companies collect data on customer transactions such as sales revenue by products, by transactions or by store locations. The transactions are often not tagged to an individual customer, meaning that the retailers only know what products were purchased in which transactions, but not who made the transactions. Personal customer data are only collected if customers are part of a loyalty programme, or place advanced orders for products that require some form of identification.

Typically, the same types of data are collected for both physical store sales and online sales. Some online retailers (e.g. blogshops) require that customers sign in, or provide an email address. These retailers are able to identify specific customers.

6.2 Data Analytics

The types of data analytics performed are in line with the types of data collected, which is primarily around improving customer experience and business operations.

Pure e-commerce companies have cohesive frameworks for data collection. They undertake some form of predictive analytics, in addition to historical data and trend analysis. Brick-and-mortar retailers, on the other hand, are generally drawing observational insights based on historical data and are only starting to become aware of analytics.

Pure e-commerce companies are also pushing to open up access to analytics to all employees. These companies are no longer reliant on a single department to provide all the analytics solutions, but rather, an open-to-all system that tries to maximise the potential of analytics.

6.2.1 Improving Customer Experience

Businesses with an online presence generate customer data from activities performed, both on the company website and, often, on other websites as well (e.g. through pixels or cookies). This wealth of data can be used by the business to improve customer experience, in many ways including:

- Data on the types of products a customer has recently viewed allows a business to predict the preferences and buying intent of the customer and start recommending similar products or services to the customer.
- Data on historical transactions allow a retailer to observe frequently purchased items and use these as “favourites” in future purchases.
- Data on the products that a customer has added to his cart, but did not checkout, allows a business to use dynamic advertising to prompt the customer of the uncompleted transaction, even on other websites or applications.
- Data on the effect of previous promotional campaigns allows a business to launch similarly successful campaigns, or target different campaigns at different types of customers.
- Data on the products that are viewed most frequently by customers allows a business to rank those higher, so customers do not have to search long before finding a desired product.
- Data on the path a customer takes through a business’s website allows it to optimise the structure and flow of its website to ensure a smooth customer experience.

Box 13 below provides an example of how analytics is used to improve customer experience at Lazada.

Box 13: Benefits of Data Analytics Solutions at Lazada

Lazada is a large pure e-commerce company, providing an online retail platform for shoppers to buy goods sold either by itself or through third party merchants.

Lazada was dealing with an increasing amount of data generated by shoppers and merchants. Initially, it was relying on manual reporting and analysis to extract information, but quickly realised the need for a more integrated platform to automate the process to gain more meaningful insights.

Lazada implemented Qlik's data visualisation tool to keep up-to-date with the massive amounts of data being generated daily.

The access to these insights and ability to perform analyses were not just limited to a single division (e.g. analytics), but to most employees. It found that, through having such a system:

- Lazada's marketing and operational department is able to better understand consumer behaviour and gain insights on supply chain processes
- Lazada's marketing team is able to develop effective campaigns for online shoppers, and track how each campaign (e.g. the Lazada Online Revolution) performs
- Lazada is in a better position to enable sellers and merchants, helping them grow their business. For example, if a merchant is falling behind in deliveries to customers, Lazada could quickly analyse the entire supply chain and transactions to identify the issue and help the merchant develop effective solutions to ensure products are delivered on time

Sources: <http://www.qlik.com/us/company/press-room/press-releases/1121-qlik-helps-lazada-gain-data-insights-deliver-optimized-shopping-experience-consumers> and <https://www.cio-asia.com/resource/analytics/lazada-optimises-shopping-experience-with-analytics/>

Businesses are also using analytics to understand the general public perception of them. Without analytics, this would have involved the manual review of newspapers and other forms of media and it would only have provided a very limited view of public perception. Through the use of analytics and techniques (e.g. web crawlers), mainstream and social media can be thoroughly scanned for what are being said about any company or any issue.

Box 14 below shows how Airbnb made use of such social media analytics services to understand how its brand performs.

Box 14: Understanding Brand Position through Media Monitoring Analytics

Airbnb is an online marketplace that enables people to lease or rent short-term lodging. It has a global presence but still wanted to remain localised in each of the countries it operates in. It partnered with Isentia, a media analytics company, to understand:

- How do travellers perceive their brand across different markets?
- Do travellers trust them?

Isentia collected data from customised crawling of specific websites (including Facebook, blogs, forums and news sites) to ensure a comprehensive database. After that, data had to be validated to detect if there were, for example, local slangs, sarcasm, grammatical nuances or culture variations.

Through the exercise, Airbnb was able to understand:

- Its performance and public perception in different market segments
- Customer behaviour across online platforms, allowing Airbnb to prioritise media outreach
- Maturity of different markets and receptivity to Airbnb's messaging, allowing Airbnb to better focus its marketing efforts

Source: <http://www.isentia.com.sg/case-studies/airbnb-s-testimonial>

6.2.2 Improving Business Operations

Retailers also use data to improve their business operations. Insights relating to business operations tend to be more backward looking as compared to predictive. Examples are:

- Data on delivery lead time allows a business to give a more accurate estimate of when products will arrive.
- Scale operations, by changing operations that used to be manual into automated functions (e.g. manual vetting of customer reviews to identify spam).
- Manage inventory, by predicting demand for goods to ensure sufficient stock and to prevent unnecessary stock inventory.

For businesses that have third party sellers on their platforms, an analysis of how each seller is doing is also important. Customers link their experience with these third party sellers to the platform itself, even if the only role of the platform was to facilitate the trade. In such instances, data analytics can be used to track:

- Revenue comparisons between sellers, and using analytics to predict the top sellers to focus on.
- How responsive different sellers are to customer queries and reviews.
- The ability of different sellers to meet demand and deliver goods, and use leading indicators to predict when a seller would likely fail to meet orders.
- How much the different sellers are being charged by the platform, in comparison to the revenues they are generating.

Competition Commission of Singapore

Understanding the Data and Analytics Landscape in Singapore

August 2017

Box 15 below provides an example of how RedMart used data analytics solutions to accelerate its business operations, which would not have been possible without advanced analytics solutions.

Box 15: Benefits of Data Analytics Solutions at RedMart

RedMart is an online grocery marketplace that stocks more than 16,000 products made up of tens of thousands of product items in a 100,000 square foot warehouse. It delivers groceries and household essentials to customers seven days a week, and therefore requires constant fast and updated insight into inventory and shipment data to meet customers' orders.

RedMart's business intelligence team wanted to keep track of customer behaviour data. However, RedMart had a business intelligence tool that had limited reporting capacity. It had data in silos and insights could not be generated from different data sources. The IT department had sole access to data, and different teams had to rely on it to generate insights from data.

RedMart wanted real-time insights into buying patterns and trends that could allow personalised customer experiences, optimal pricing and improve customer service. It made use of analytics solutions provided by Tableau and Amazon Redshift to drastically turnaround its data and analytics situation.

Some of the benefits from using a data analytics solution catered to its needs were:

- Merchandising and commercial teams study margins and analyse category performance for optimal product selection
- Marketing team studies the uptake of various initiatives and improve promotional offers
- Customer service team tracks the number of customer contacts for any given day or week, based on the type of contact and reason for contact

These then translated into:

15-20%	4,000	5X	60%
increase in online orders	peak orders per day	front-end website speed	cheaper website operations

Also, when employees can individually access data to answer business questions, RedMart's business intelligence team is now spending less time coordinating requests and generating reports, but focusing on other strategic efforts.

Sources: <https://www.tableau.com/solutions/customer/redmart-cuts-analytics-time-70-80-percent-tableau> and <https://aws.amazon.com/solutions/case-studies/redmart/>

6.2.3 Other Retail Companies

Other retail companies, especially the brick-and-mortar ones, make very limited use of data analytics. Data that are collected are mainly used for BI, and targeted marketing

campaigns can only be directed at customers who leave personal details with the retailers.

6.3 Data Sharing

Customer data are not shared between companies. Pure e-commerce companies thrive on the customer data they possess, including customer preferences and buying behaviour. It is the generation of these insights from analytics that provides pure e-commerce companies with a comparative advantage, when they are used to target advertisements and cater solutions to customers.

6.4 Data Storage

Retail companies (both pure e-commerce and brick-and-mortar) tend to store data physically, rather than on the cloud. The concerns with cloud storage primarily relate to cloud providers having control over the business's data, as well as the difficulty in retrieving data if, for example, the business wants to switch providers.

Data storage is distinct from website hosting, and pure e-commerce companies that do not store data on the cloud can still use the cloud for hosting its business.

6.5 Data Monetisation

Among pure e-commerce companies, data monetisation occurs through targeted marketing. Just like how they are not shared, customer data are also never sold. Putting aside concerns about data privacy, the value that these companies are able to derive from targeted marketing far exceeds the value that the data can sell at currently.

The monetisation process is shown in Figure 6.2 below.

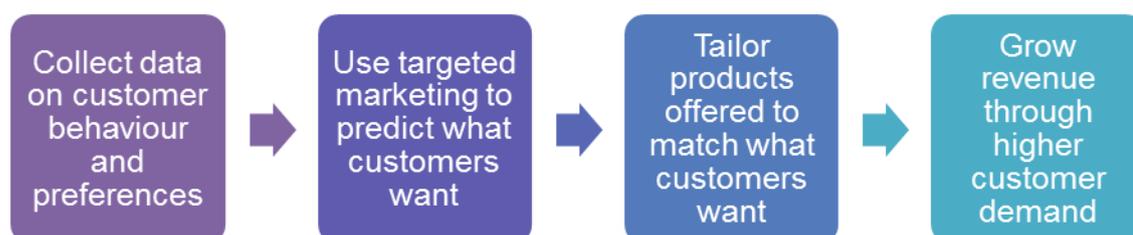


Figure 6.2 Business Process for Monetising Customer Data

Retail companies that have limited data collection and use of analytics have limited opportunities for data monetisation.

6.6 Impediments to Greater Use of Data Analytics

Similar to the impediments faced in other sectors, the lack of skilled human resource is a significant impediment. Pure e-commerce companies, in particular, require awareness

and knowledge of data and analytics by employees across the entire organisations, as they push to allow open access for all.

Box 16 below is an example of how data education is done at Airbnb.

Box 16: Airbnb’s Data Education

The Data Science team in Airbnb strongly believed in empowering all employees in the company to understand and work with data. *“In order to inform every decision with data, it wouldn’t be possible to have a data scientist in every room – we need to scale our skillset.”*

Airbnb took the task of ensuring all employees made use of data analytics very seriously. It tracked this by using a metric of weekly active users of its data platform as a proxy of how “data informed” they were. At the beginning of Q3 2016, this number was only at 30 per cent, significantly lower than its peers like Facebook and Dropbox.

An internal assessment found that the largest bottleneck to scaling data informed decisions was the lack of data education for users. Airbnb therefore created the Data University, comprising over 30 classes covering an array of different topics. Through this, Airbnb grew to have a “data informed” score of 45 per cent in 2017.

Source: <https://medium.com/airbnb-engineering/how-airbnb-democratizes-data-science-with-data-university-3eccc71e073a>

6.7 Implications of Data Protection Laws

Strict data protection laws mean that data that are collected have to be anonymised immediately. The analysis of customer segments and purchasing patterns do not rely on individual customer data but aggregated data.

6.8 Summary of Data Analytics Maturity

The data and analytics practices and maturities between pure e-commerce and brick-and-mortar retailers vary significantly. Brick-and-mortar retailers are only becoming aware of data and analytics, whereas pure e-commerce retailers already have cohesive data collection frameworks and are using predictive analytics to drive some aspects of business operations. (Figure 6.3)

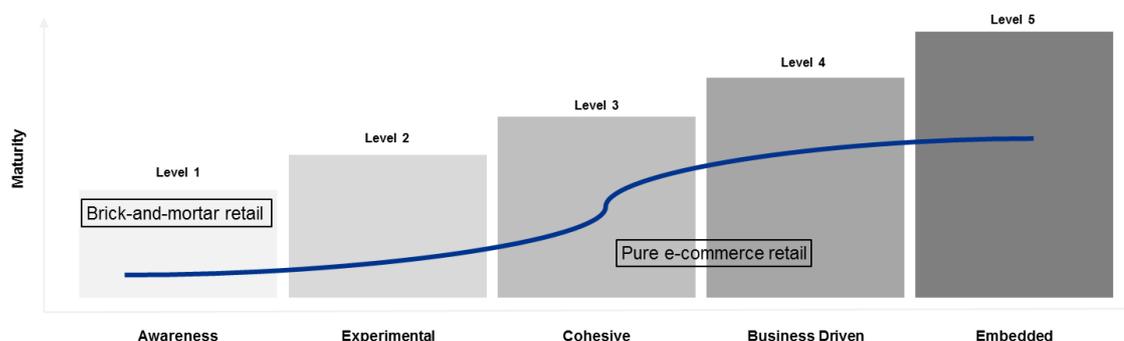


Figure 6.3 Maturity Map Assessment of the Consumer Retail Sector

7 Land Transport Sector

The land transport sector is very diverse, and ranges from private ride booking companies³⁷ (e.g. Uber, Grab) to public transport operators and private bus hiring companies. Similarly, the analytics capabilities range from highly advanced to very basic. The Land Transport Authority (LTA) drives the analytics initiatives among public bus operators, and also facilitates data sharing in the sector.

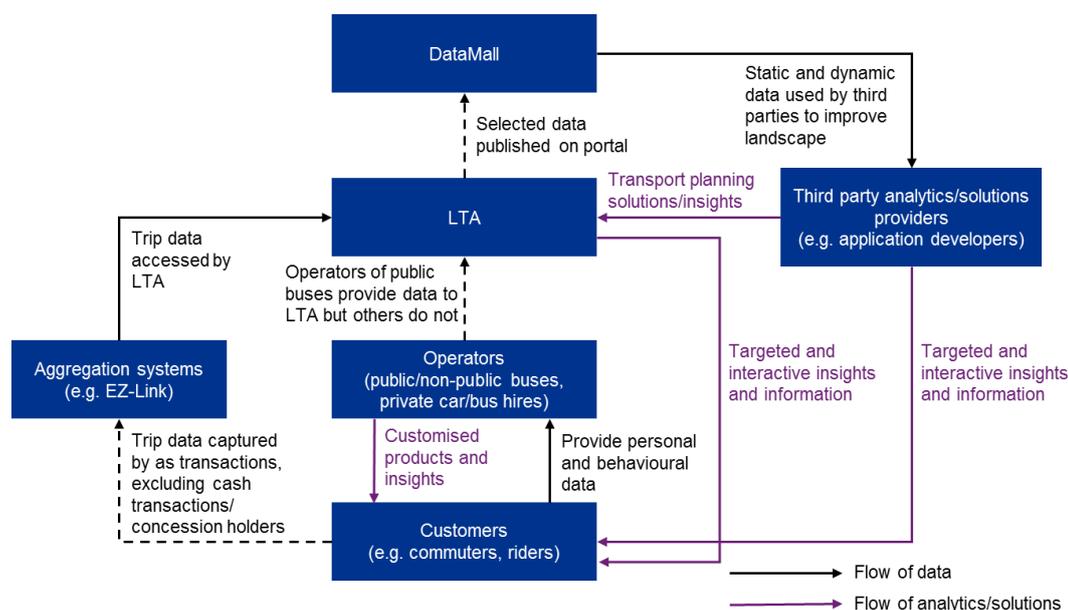


Figure 7.1 Data and Analytics Landscape in the Land Transport Sector

As shown in Figure 7.1, the main players in this sector are:

- **LTA** – the operator of DataMall, an online portal housing transport data, which is also in charge of the overall public transport planning process.
- **Operators** – ranging from operators of public buses to taxis and operators of private hire bus and ride booking services.
- **Aggregation systems** – which facilitate e-payments and collect data on commuter journeys.
- **Customers** – who are either commuters on regulated services or customers on non-regulated services.
- **Third party application developers** – which have arisen in recent years to provide value-added services to the wealth of data made publicly available by LTA.

³⁷ By definition, these companies refer to themselves as “technology companies”. However, for ease of discussion, this report refers to them as private ride booking companies.

7.1 Data Collection

In the public transport sector, the types of data collected primarily relate to customer travel patterns and business operations.

7.1.1 Data Collected by LTA

LTA collects a lot of commuter data. Data from aggregation systems are able to map most commuter journeys reliably, as the system tracks where each commuter gets on or off a bus/MRT train.

LTA is also putting in place a Common Fleet Management System (CFMS), which is a centralised system across all bus operators that does operations control, fleet management, passenger information and business management. There are two parts to the system:

1. **An Intelligent Bus Management System** – comprising computer systems installed within the bus depot.
2. **In-Vehicle System** – a touch-screen tablet mounted on every bus and associated location systems.

Aside from allowing bus operators to control their fleet and manage operations, LTA will also have access to bus performance data.

7.1.2 Data Collected by Providers of Private Ride Booking Services

Private ride booking service providers such as Uber and Grab collect customer data and information through two main ways:

1. **Provided by customers** when they sign up as members. Such data are personal, including email address, mobile number and payment details.
2. **Collected from customers** when they conclude transactions (i.e. take trips) using the application. These are behavioural, relating in particular to travel patterns (origin, destination, frequency, time of use) and preferences (type of car).

These companies do not collect a lot of personal customer data (e.g. age group, gender) as travel pattern data have more value.

7.1.3 Data Collected by Providers of Private Bus Hire Services

Private bus operators collect limited data on business operations and customers, due to their B2B business model. Contracts are often negotiated for a fixed capacity or passengers at fixed schedules, and payments do not depend on the occupancy or capacity utilisation of the buses. Depending on the type of service provided, data collected are:

- **Scheduled bus routes** – Bus location data are collected, depending on clients' needs.

- **School bus routes** – Operators of school buses, especially those ferrying students from international schools, collect data on when students board or get off the bus, and notify parents of these.
- **Shuttle bus routes** – Data on how many people are on board each trip are collected through people-counting machines, which can be used to optimise future trips.

7.1.4 Data Collected through Crowd-Sourcing

Beeline is an initiative by LTA and the Government Technology Agency of Singapore (GovTech) that allows commuters to “crowd-start”. Crowd-starting is a form of crowd-sourcing, which refers to the data collection process that involves a large crowd of people instead of a single provider of data.

Commuters indicate, through the Beeline application, their preferred commuting routes and demand on those routes and bus operators respond to these by providing shuttle bus services. Crowd-starter routes only activate after a minimum demand threshold is met. As of March 2017, there are two such crowd-starter routes have been activated and sixty other crowd-starter routes have been launched by Beeline.³⁸ Through this, commuters are able to participate in the route designing process, and bus operators are able to tailor demand to meet these needs.

7.2 Data Sharing by LTA

Selected data collected by LTA are published on DataMall, an online portal accessible by the general public. The aim of the portal is to promote citizen co-creation of innovative and inclusive transport solutions.

This portal serves as the main source of data in the industry, and industry players generally do not undertake their own data collection process if the dataset is already available on DataMall.

There are two main types of data shared on DataMall:³⁹

1. **Static data** – ridership, transport infrastructure, car fleet.
2. **Dynamic data** – bus location (most popular dataset), taxi availability.

Dynamic data are shared with users through APIs, which update bus locations in real time. API access is obtained through an application to LTA, and as such, LTA has oversight of the users of the data.

The main users of dynamic data on DataMall are application developers. Of these, most of them tap into bus location services, which is discussed further in Section 7.4.

³⁸ <https://www.tech.gov.sg/-/media/GovTech/Media-Room/Media-Releases/2017/0302/Beeline-Factsheet.pdf>

³⁹ See <https://www.mytransport.sg/content/mytransport/home/dataMall.html> for a list of data available.

7.3 Data Sharing and Collaboration between Industry Players

Uber is beta-testing Uber Movement, a new initiative that provides users with access to anonymised historical travel data.⁴⁰ Users are able to get an estimate of travel times between two points in a city, based on day of the week and time. Uber Movement was designed with urban planning in mind, and aims to help policy makers understand the impact of road improvements, new transit lines as well as future infrastructure investments.

The sector also has other forms of collaborations. For example, in March 2017, Grab partnered with five taxi companies (Premier, Prime, Trans-cab, HDT and SMRT) to come up with a dynamic fare pricing service – JustGrab. Grab contributes its analytics capabilities in dynamic fare pricing, which varies according to real-time demand and supply for taxis while the taxi companies provide the asset to fulfil the taxi rides.

Before the partnership, taxi fare structures were static and did not take into account real-time passenger demand and driver supply. This often meant that passengers paid surcharges even when there were many available taxis within the vicinity.⁴¹ It is unlikely that taxi operators, especially the smaller companies, would have had the resources to develop a similar system on their own. For example, ComfortDelGro's option of flat fares do not have surge variables, and only take into account existing surcharges.⁴²

7.4 Data Analytics

Data analytics is undertaken in a variety of ways, by both LTA and industry players.

7.4.1 Bus Fleet Management and Schedule Adherence

Bus operators have adopted the use of data analytics in their business operations, primarily due to LTA's initiative with the CFMS. The main features of the CFMS are:

- **GPS bus tracking** which matches current bus position to indicated route for that service.
- **Expected bus arrival time** at each bus stop, and subsequently, the excess wait time or adherence to scheduled arrival time for each bus at each bus stop.
- **Distance ahead from the next bus** on the same service, including load of the next bus (seats available, standing space available, limited standing space available).
- **Communication** between bus drivers and the operators' respective control centres via voice or data.

The CFMS helps bus drivers adhere to their route schedules and ensure that there is no "clustering" of buses at bus stops.

⁴⁰ See: <https://movement.uber.com/cities>

⁴¹ <https://www.grab.com/sg/press/consumers-drivers/smrt-taxis-signs-grab-exclusive-app-partner-collaborates-dynamic-fixed-fares-taxi-rides/>

⁴² <http://www.straitstimes.com/singapore/transport/comforts-flat-fare-option-gets-thumbs-up>

7.4.2 Public transport planning

LTA has set out the Smart Mobility 2030 Master Plan, where an Intelligent Transport System will be in place to enhance commuters' travelling experiences in Singapore. The possession and ability to interpret data are critical to implementing the system. The Master Plan sets out three key strategies, and four areas of focus to achieve this.



Figure 7.2 Focal Areas and Key Strategies of LTA's Smart Mobility 2030 Master Plan
 Source: <https://www.lta.gov.sg/content/ltaweb/en/roads-and-motoring/managing-traffic-and-congestion/intelligent-transport-systems/SmartMobility2030.html>

LTA has partnered with IBM to implement the Fusion AnalyticS for public Transport Emergency Response (FASTER), which combines video data at selected MRT stations from SMRT and telecommunications travel data from StarHub for crowd management across the public transport network in Singapore.⁴³ FASTER will help LTA visualise commuting patterns to improve transport planning, triggering early alerts of crowd surges and transport incidents.

LTA also uses data visualisation techniques to observe travel patterns on bus routes, through EZ-Link transactions, to map out the “fingerprint” of bus routes in Singapore. Bus stops are plotted as dots in a semi-circle, and trips between stops shown as lines between the dots. The thicker and higher the lines, the higher the number of passengers that travelled between those two stops. This is shown in Figure 7.3 below.

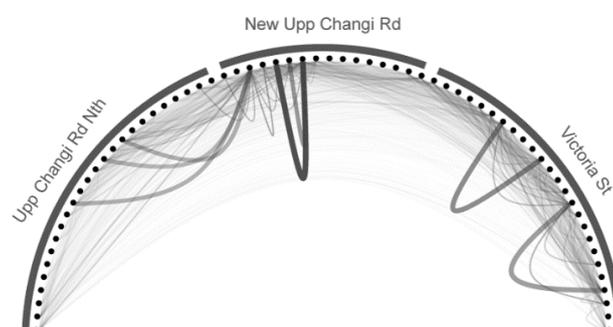


Figure 7.3 Fingerprint of Bus Route Number 2
 Source: <https://blog.data.gov.sg/fingerprint-of-a-bus-route-73e5be53dcf0>

⁴³ <https://www.lta.gov.sg/apps/news/page.aspx?c=2&id=407a5053-0345-40f5-8d64-51fb31bfb2a0>

Competition Commission of Singapore

Understanding the Data and Analytics Landscape in Singapore

August 2017

The fingerprints can be produced for different services, different times of day or different days of the week. Some conclusions that were drawn from the exercise were:

- Most bus routes are busier during weekdays, although some bus services have no discernible differences in ridership between different days of the week.
- Typically, weekday morning peaks are earlier (6 a.m. to 9 a.m.) than weekend morning peaks (after 9 a.m.).
- There is higher utilisation of the bus closer to bus interchanges and MRT stations, with relatively few trips spanning the entire bus route.

Based on these findings, the exercise concluded that cutting bus routes could reap greater efficiency, especially since commuters generally do not stay on the bus for the entire route. However, transport planning would have to take into account the challenges of managing the impact to existing commuters and the availability of suitable terminating points.

Box 17 below shows an example of how traffic data can be combined with other data on commuters to further increase the effectiveness of public transport planning policies.

Box 17: Simulation of Human Mobility for Sustainable City Planning

To manage the morning peak hour traffic situation, LTA previously experimented with offering free early morning train rides to MRT stations in the city area before 7.45am on weekdays. The results of the free rides have yet to be quantified.

DataSparks, Singtel's data analytics subsidiary, is proposing that data on mobile subscribers' locations and customer segmentation can be used to analyse the effectiveness of such schemes. Where transport data could only identify the overall percentage of commuters who switched to the free rides, mobile data could complement that to determine the breakdown between customer segments and profiles.

Source: <https://datasparkanalytics.com/insight/can-data-science-help-build-better-public-transport/>

7.4.3 Making Public Transport More Accessible to the General Public

In line with LTA's third key strategy in its Smart Mobility 2030 Master Plan to establish close partnerships and co-creation, third party application developers have also contributed to the availability of data analytics solutions in the public transport sector. A range of developers have created applications that are able to predict when the next bus arrives, relying on data from DataMall. Examples of such applications are "SG Buses", "Singabus", "SG Buses Delight", "SG BusLeh" and "MyTransport Singapore". Commuters are able to select the bus stop and bus service, and get data on when the next bus is arriving.

One analytic solution that differentiates itself is "Bus Uncle", which uses a chatbot on a Facebook profile to interact with its users. Unlike the applications that provide information on all bus arrival timings at a bus stop, Bus Uncle only gives the commuter the information he needs (i.e. only for that bus service). It is also localised to the Singapore context, incorporating the use of Singlish and other local phrases, and has a "personality"

Competition Commission of Singapore

Understanding the Data and Analytics Landscape in Singapore

August 2017

that is supposed to resemble that of a typical bus driver. The page currently has more than 26,000 followers and 'likes', indicating its huge success (Figure 7.4).



Figure 7.4 Bus Uncle, a Third Party Application Relying on Bus Location Data on DataMall

Source: Bus Uncle's Facebook profile

The page also uses crowd-sourcing to keep itself up to date. For example, if users know of a new bus route, they are encouraged to raise this to the Bus Uncle page so that new data can be incorporated.

7.4.4 Real-time Demand and Supply Matching

Private ride booking companies (e.g. Uber and Grab) are the most advanced in their use of data analytics in the sector. Grab, for example, utilises analytic tools that are able to assess demand and supply in Singapore in each 1km by 1km area to match drivers with riders. The difference between demand and supply then affect the price quoted to the rider.

Grab builds its operations on the Amazon Web Services platform, using a number of features such as:⁴⁴

- **Elastic Load Balancing** – which is able to take into account peak hours and scaled up demand.⁴⁵
- **Analytic services** – which allows Grab to react to situations of high demand, such as on rainy days or when an MRT train breaks down.

Other than that, Grab also uses analytics to gauge the odds of a driver bidding for a job. If a driver gets a notification of demand at a certain location, but does not go there, the behaviour patterns are captured as data which are then used to personalise the types of notifications the driver receives in the future.

7.5 Data and Analytics Monetisation

The monetisation of data and analytics is possible due to the availability of data through LTA's efforts.

⁴⁴ <http://www.tnp.sg/news/singapore/riding-data-better-efficiency>

⁴⁵ <https://aws.amazon.com/elasticloadbalancing/>

7.5.1 Monetisation by Providers of Raw Data

The data that LTA publishes on the DataMall is free of charge; there are no costs to using it, other than the application to LTA for API access. While there are costs involved in collecting and publishing the data, there are positive externalities to the use of data which cannot be quantified. The third party applications, in particular, have helped make the commuter journey more pleasant and public transport, as a whole, more manageable.

7.5.2 Monetisation by Users of Data

Third party applications that build on LTA's bus location data do not currently charge for downloads and use. Monetisation therefore typically happens indirectly, through advertisement revenue. Popular applications have higher traffic and as such, are able to generate higher advertisement revenue.

Private bus operators monetise the crowd-sourced data provided through Beeline, by providing services that meet commuters' demand.

7.6 Data Storage

Data storage practices vary across companies in this sector. The private ride booking companies readily utilise cloud storage while other companies do not yet.

7.7 Implications of Data Protection Laws

Similar to the practices in the other sectors, data collected by LTA and other service providers are aggregated and anonymised. Data cannot lead to the identification of individuals.

7.8 Summary of Data Analytics Maturity

The transport sector's use of data and analytics is assessed as experimental, mostly through LTA's initiatives. With the exception of private ride booking companies and LTA's initiatives, most companies do not have proper frameworks for data collection and are just experimenting with analytics. (Figure 7.5)

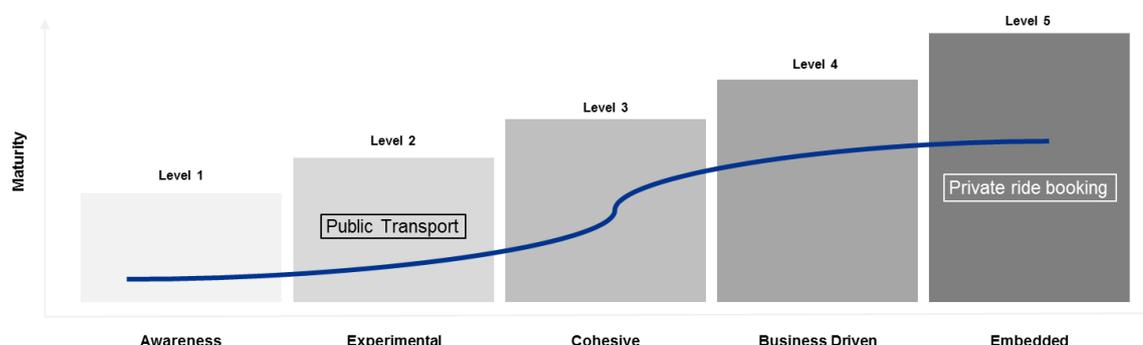


Figure 7.5 Maturity Map Assessment of the Public Transport Sector

8 Logistics Sector

The logistics sector has traditionally been slower in applying data analytics to its business operations, although this is expected to change with initiatives from selected leading players. Singapore's dense road network, small size and relative proximity of delivery destinations have allowed traditional business models to survive, a phenomenon that is prevalent globally.⁴⁶

The Infocomm Media Development Authority (IMDA) has also made significant efforts to improve productivity through technology and analytics in this sector, and is focused on maintaining a level playing field for all logistic service providers.

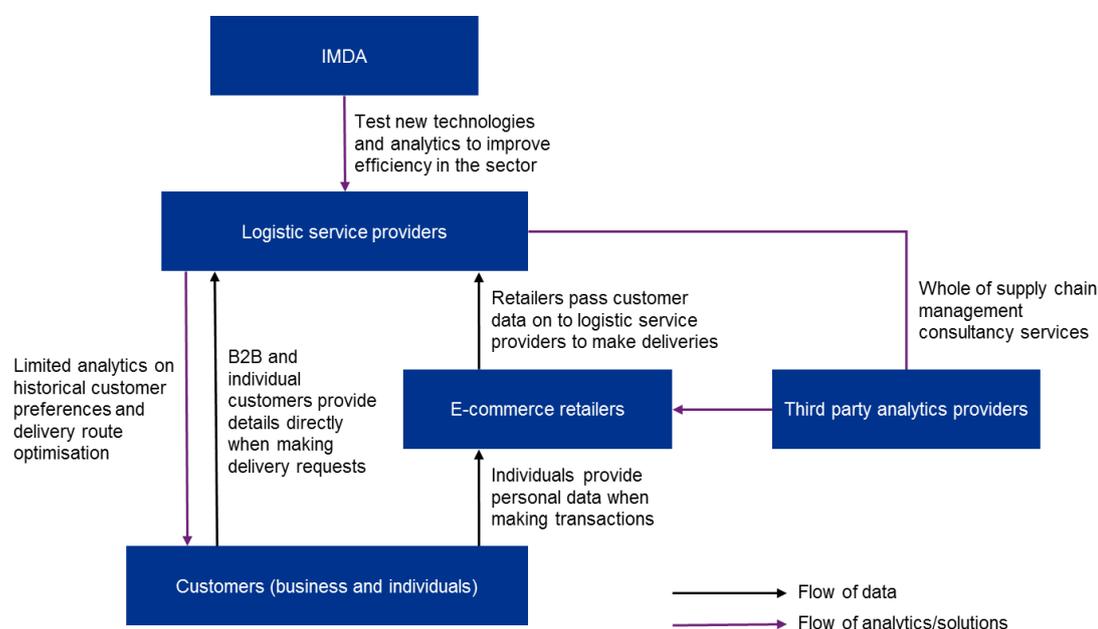


Figure 8.1 Data and Analytics Landscape in the Logistics Sector

As shown in Figure 8.1, the main players in this sector are:

- **Logistic service providers** – which provide the delivery network and collect data on customers' transactions. There are two main types of providers – those that deliver between businesses (B2B) and those that deliver from businesses to individuals (B2C).
- **E-commerce retailers** – which generate delivery demand through end customers' purchase transactions.
- **Individual customers** – who use delivery services, either directly or indirectly through e-commerce transactions, and similarly provide personal data to logistic service providers directly or through retailers.

⁴⁶ World Economic Forum (January 2016) *Digital Transformation of Industries – Logistics Industry*

- **Business customers** – which use delivery services through longer-term agreements with logistic service providers.
- **Third party analytics providers** – which usually provide consultancy services to retailers in managing their supply chains, and are usually provider-independent. The limited amount of analytics undertaken by logistic service providers means that there is little collaboration with third party providers for internal analytics.
- **IMDA** – which facilitates the use of technology and analytics in the sector through its initiatives under the Urban Logistics technology roadmap 2020.

8.1 Data Collection

Logistic service providers collect customer data to fulfil their delivery orders, and shipment data to optimise delivery options and routes.

8.1.1 Collection of Customer Data

Logistic service providers collect data on customers when they perform deliveries. The types of data collected include parcel origin, sender personal details, recipient address, delivery method and option (e.g. express, standard) and special delivery instructions.

Deliveries originate through two ways; they can be directly requested by customers (both businesses and individuals) or they can be requested through an e-commerce retailer as part of its order fulfilment. Customers that select the delivery option after an online purchase often do not have sight of the logistics company that will fulfil the order. Customers only choose between delivery speed and cost.

A distinction has to be made between sender and recipient personal data:

- Sender data are treated like how other companies treat customer data. Logistics service providers can generate insights from purchase patterns and send out promotional advertisements, subject to customer consent.
- Recipient data, however, are transient data. Logistic service providers have access to that data during the delivery, and can only use the data to fulfil the delivery. Once the delivery is complete, such data can no longer be used to generate any insights or for marketing purposes.

8.1.2 Collection of Shipment Data

Logistic service providers also require data on shipment routes, especially for international deliveries. Shipping cargo from Singapore to New York, for example, can go through numerous routes with many permutations of transport options (e.g. air, land, sea). Senders only require that cargo reach their destination within a certain amount of time; they do not have a choice on the route it takes. Logistic service providers are therefore able to choose the route that incurs the least cost.

Shipment data are not structured, and there is no central database to obtain data. A lot of the data, or the knowledge of data, comes through experience with having used the carrier previously, or through word of mouth.

8.2 Data Analytics

Data analytics has traditionally been limited, but this is changing. The rise of e-commerce has generated partnerships with logistics service providers to provide better delivery services and lead times. IMDA is piloting initiatives for the sector, but it is up to industry players to take up the initiatives once they are ready for the market.

8.2.1 IMDA's Urban Logistics Technology Roadmap

IMDA has set out the Urban Logistics technology roadmap for 2020, which tests and implements new technologies to enable greater supply chain efficiencies in the logistics sector in Singapore. In essence, the roadmap involves the sharing of logistics resources such as vehicles, drivers and warehouses to coordinate delivery schedules and improve overall effectiveness in the sector.

In October 2015, the Government announced the allocation of \$20 million to the transformation of domestic logistics, and the roadmap was drafted to test how different stages of the logistics process can benefit from technology. The solutions, systems and processes developed under the roadmap has to be interoperable and open to industry players to adopt or adapt. This is to ensure that a level playing field is set for small and medium enterprises in the logistics sector.

One of the first areas for improvement identified by IMDA was the last-mile delivery system within shopping malls. Shopping malls did not have a proper system that optimised delivery schedules, which resulted in:

- Congestion often occurring at the loading and unloading bays as logistic service providers clustered their delivery times.
- Delivery trucks and vans spending a lot of time waiting at the loading and unloading bay instead of performing more deliveries.
- Individual retailers separately retrieving cargo, even if they were situated near each other and had similar delivery times.

The In-Mall Distribution system helped to solve these problems, as described in Box 18.

Box 18: Improving Delivery Systems within Shopping Malls

The In-Mall Distribution system was introduced in June 2016 in Tampines Mall and in September 2016 in Bedok Mall. There are plans to roll this out to four more malls in 2017 – IMM Building, JCube, Westgate and Bukit Panjang Plaza.

The system relies on various technologies, such as a dock scheduler and queue management system, which allow logistics companies to book delivery time slots. A logistics operator is stationed in the mall to receive and consolidate deliveries from suppliers. The operator then makes the last-mile delivery to retailers within the mall. The operator can offer value-added services such as unloading bay facility management, security pass management, concierge services or meal deliveries.

Stakeholders at Tampines Mall have benefited from the programme through:

Competition Commission of Singapore

Understanding the Data and Analytics Landscape in Singapore

August 2017

- Retailers saved time and resource from stock transfer
- Traditional brick-and-mortar retailers expanded into the e-commerce space through the operator's outbound delivery services to homes or offices
- Reduced traffic congestion at the unloading bay due to more spread out delivery timings
- Logistic service providers reduced their waiting time from 24 minutes to seven minutes, and were able to better utilise and allocate assets

Source: <https://www.imda.gov.sg/about/newsroom/media-releases/2016/launch-of-urban-logistics-technology-roadmap>

Another example of inefficiency in the logistics supply chain is multiple trucks making deliveries, all with less than full loads. This is evidenced when different logistic service providers make deliveries to the same destinations, with no knowledge of each other's deliveries, or ways to share resources and combine deliveries. In such instances, assets are inefficiently used, resulting in more vehicles on the road than necessary. IMDA's Offsite Consolidation Centre is an initiative that tries to address this problem (Box 19).

Box 19: Improving Truck Utilisation by Sharing Resources

IMDA implemented the Offsite Consolidation Centre to allow transported cargo to be consolidated and sorted, before they are delivered to their final destinations. Cargo is re-loaded and then delivered to their intended destinations on a single truck within the same day. Truck load utilisation is improved, and the number of trucks going to the same destination is reduced.

Advanced robotics can also be deployed to help with sorting cargo, while fleet optimisation solutions can enable smooth, tracked and optimised deliveries.

Source: <https://www.imda.gov.sg/about/newsroom/media-releases/2016/launch-of-urban-logistics-technology-roadmap>

8.2.2 Improving Business Operations

Logistic service providers are increasingly starting to adopt more data analytics in a sector that did not traditionally rely on these tools. Examples of some of the uses of analytics to improve business operations, although on a limited scale by selected logistics service providers, are:

- Warehouse automation that makes use of autonomous vehicles and devices that have optimised routes and flow of materials.
- Collaboration with pure e-commerce companies to predict customer demand and deploy stock pre-emptively to a nearby location to reduce delivery lead times.
- In-vehicle sensors to monitor fleet in real-time. Data on drivers' driving patterns (e.g. speed, braking habits, vehicle idle time) are monitored, primarily for safety and to reduce insurance premiums on vehicles, rather than for route and delivery schedule optimisation.

Competition Commission of Singapore

Understanding the Data and Analytics Landscape in Singapore

August 2017

- Mini-routes are identified within larger delivery routes, and historical data used to provide better estimates of delivery lead time once the delivery man starts on the mini-route. This provides better estimation of delivery times to customers, and to reduce the chances of failed deliveries from customers not being at home.
- Historical data on customer delivery preferences (e.g. enter through the back door) are stored and referenced for future deliveries.
- Real-time tracking of parcels to provide delivery status updates to customers. (see Box 20 below)

Box 20: Real-Time Updates on Parcel Delivery

SingPost recently launched its SmartPost delivery system, which allows real-time tracking of parcels and sends delivery status updates to customers along each step of the delivery process. SmartPost makes use of near-field communication technology, radio frequency identification, digital imaging and electronic notifications.

More than 1,000 postal delivery workers across the island will be equipped with smartphones loaded with a customised postal delivery app.

SmartPost works across SingPost's entire postal operations, from collection to sorting, last-mile delivery and quality assurance. It is expected to elevate SingPost's operational capabilities and operating efficiency.

Source: <http://www.todayonline.com/singapore/singpost-unveils-smartpost-boost-service-quality>

The logistics sector is not immune to disruptive technology. The limited use of analytics by traditional logistic service providers and the large efficiency improvements that could arise from this use represents a market arbitrage opportunity. Ninja Van is an example of a company that actively uses analytics to optimise delivery schedules and manage its fleet, in contrast to the reliance on human experience in other logistic service providers. Box 21 below describes this.

Box 21: Ninja Van Disrupts the Logistics Sector

Ninja Van is a start-up that tried to reinvent the logistics sector using data analytics. Ninja Van was one of the first to digitally scan parcels, while the rest of the sector was still using outdated mail-sorting technology such as manual sorting. It examined a parcel's entire life cycle, from the point it is picked up to the point it is delivered.

Ninja Van actively uses algorithms to calculate the best routes a driver should take, or which van should be used to deliver a parcel. This not only allowed drivers to deliver parcels more quickly, but also helps save fuel costs.

Source: <http://www.channelnewsasia.com/news/singapore/how-ninja-van-s-smart-and-ruthlessly-aggressive-co-founder-hooke-8601638>

8.2.3 Potential Use of Predictive Analytics to Forecast Demand

The ability to forecast demand for delivery services is an important value-added service that logistics service providers can incorporate into their operations. At present, logistic service providers observe historical trends in delivery demand (e.g. higher demand towards the end of the month), and only a few providers are experimenting with forecasting demand.

Demand forecasting is beneficial for both the logistic service provider and the customer:

- Logistic service providers minimise the strain on its resources in coping with demand that exceeds normal capacity.
- Customers reduce delivery costs by not paying last-minute and peak rates.
- Errors from rushed orders are avoided, both from the logistic service provider and the customer.

8.3 Data Sharing

There is no sharing of data between logistic service providers. Data sharing is limited as delivery data are often confidential (e.g. type of cargo, who the cargo belongs to).

8.4 Data Monetisation

The limited use of analytics means that there is also limited scope for monetisation. Companies with more advanced analytics systems (e.g. Ninja Van) have tried to monetise their capabilities by offering delivery services that are different from those offered by traditional providers.

8.5 Data Storage

Generally, data (especially sensitive customer data) are stored physically.

Concerns with security and the confidentiality of customer data have limited the use of cloud storage, although some less sensitive data are being put on the cloud. Costs of migrating data have to also be considered if cloud storage is being used.

8.6 Impediments to Greater Use of Data Analytics

Similar to the challenges faced in other sectors, the high costs of analytics systems and the availability of skilled labour are also problems faced by logistic service providers. The fragmented market with multiple providers mean that competition is largely based on costs, creating further disincentives for investments in analytics solutions.

8.7 Implications of Data Protection Laws

Data protection and privacy are ensured as logistic service providers are not allowed to use the transient individual recipient data they receive for marketing purposes.

8.8 Summary of Data Analytics Maturity

The logistics sector's use of data and analytics is assessed to be at the experimental stage. On the whole, while the sector has traditionally lagged behind other sectors in the use of analytics, the rise of e-commerce has stimulated the use of data analytics to provide better services and lead times. Selected companies are experimenting with analytics and IMDA's efforts are significant in encouraging greater use of analytics. (Figure 8.2)

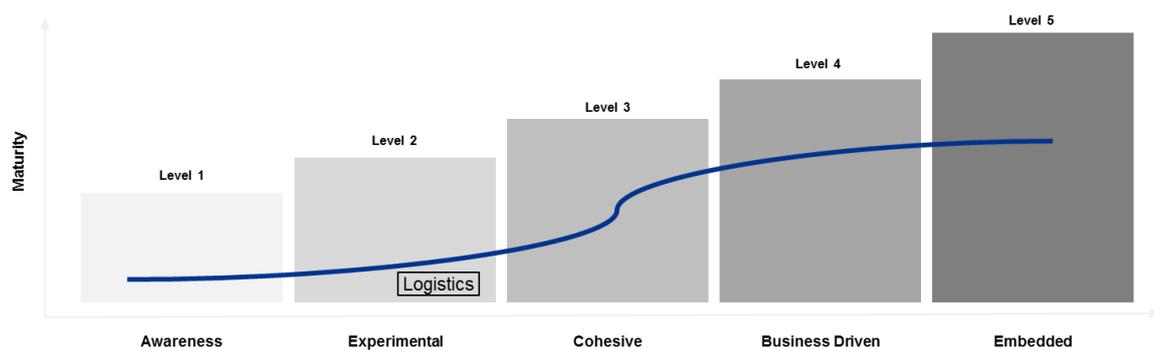


Figure 8.2 Maturity Map Assessment of the Logistics Sector

9 Conclusion

This section sets out the conclusions of the study.

9.1 Comparing Data Analytics Capabilities in the Six Sectors

In conclusion, the six sectors studied have very different maturities in relation to data collection and use of analytics. The digital media sector and the private ride booking companies are the most advanced, while most sectors are only at the first two stages of the analytics maturity map.

This situation is common across organisations and similar sectors globally, and is not specific to Singapore. The global push towards collecting and using more data aims to move organisations and sectors away from the first two stages of maturity. The Singapore Government has been making significant strides in growing and enhancing the analytics capabilities in the different sectors, as described through the report.

A number of sectors, and companies within each sector, mistake analytics with BI. Drawing business and customer insights from historical data are prevalent practices, as well as other IT systems around inventory planning and allocation. As described, analytics refers to forward looking techniques with predictive abilities. True analytics, and the real value of it, are only used and realised by a handful of companies.

The maturity assessment of analytics capabilities in the six sectors are summarised in Figure 9.1 below. It is important to note that the assessments are overall representations of the companies in that sector. Within any sector, there is a dispersion of analytics capabilities between companies.

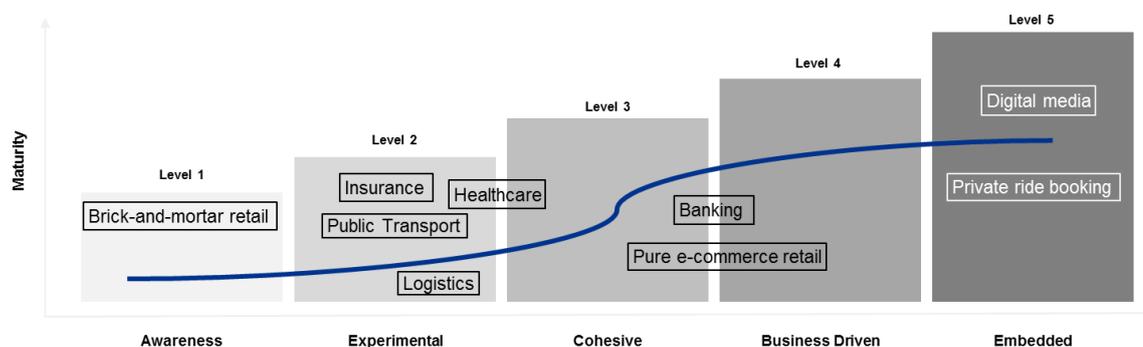


Figure 9.1 Overall Maturity Map Assessment

Two key issues drive the use of and overall capabilities data analytics in sectors:

- **Government initiatives** – the efforts of Government agencies, such as LTA, MAS and IMDA, in the respective sectors have helped to ease the high costs associated with implementing and setting up analytics infrastructure and facilitated data sharing.
- **Presence of international companies** with very advanced analytics capabilities – the presence of international companies such as the private ride booking companies (e.g. Uber, Grab) and the pure e-commerce companies (e.g. Lazada) in the respective sectors in Singapore have disrupted business operations

globally, and stimulated the development of analytics capabilities in those sectors in Singapore.

9.2 Data Monetisation

Businesses guard customer data strictly, and do not monetise them by directly selling data. Monetisation occurs indirectly, through the insights on customer preferences, behaviour and buying intent that can be drawn from the data.

Where the Government is involved in data sharing or testing and implementing analytics solutions, there is no direct monetisation. Its objectives are often to facilitate industry development, increase productivity or improve consumer experience.

External data aggregators do monetise data, but often after analysis and segmentation have been done on both customer and industry data.

9.3 Data Storage

Cloud storage services have been identified as the way forward. However, many businesses are still wary of such technology, in part due to the apparent loss of control over the data, as well as the over-reliance on one storage provider. Regulators have so far not prohibited the use of cloud storage services in Singapore.

9.4 Impediments to Greater Use of Data Analytics

Across sectors, the lack of skilled labour has been identified as the main impediment to greater use of data analytics. In particular, it is the lack of experienced labour with industry experience that is impeding the growth of analytics, as graduates are not provided with sufficient guidance.

Other impediments relate to the costs of setting up infrastructure and IT systems to support the use of data analytics. Individual companies do not immediately see the benefits of such investments, and often, a lot of the initiatives are Government-led.

There are also requests for there to be greater transparency and proper frameworks setting out the types of data that can be shared, especially those relating to personal customer details. This would facilitate greater data sharing.

9.5 Implications of Data Protection Laws

The PDPA protects data that can be used to identify individuals. Before collecting, using or disclosing such data, organisations have to ensure that customer consent is obtained for that specific and reasonable purpose.

The implication of the data protection law means that businesses are very cautious in the handling of customer personal data. Such data are anonymised, encrypted and aggregated when they are used to make customer insights, and only selected employees in an organisation have access to individual customer data.



Competition Commission of Singapore

Understanding the Data and Analytics Landscape in Singapore

August 2017

The finance sector, for example, also has additional rules around customer data protection (through the Banking Act). Strict customer data rules have meant that there is often less than optimal sharing of customer data, even within different departments in a bank. Across sectors, frameworks could be introduced to provide greater clarity on the types of data that can or cannot be shared, to improve data sharing practices.



A Stakeholder Survey Questionnaire

Introduction

Thank you for agreeing to be part of this survey into understanding the Data and Analytics Landscape in Singapore. KPMG is engaged by the Competition Commission of Singapore ("CCS") to perform this study.

One of the recommendations from the Committee for the Future Economy is the building of strong digital capabilities in Singapore. This involves promoting the adoption of digital technologies across all sectors of the economy. As data becomes increasingly ubiquitous in the digital economy, CCS is interested to understand the current industry landscape and companies' data usage and sharing practices in Singapore. This survey forms part of CCS's broader study on the implications of the prevalence of data and data analytics on competition policy and law, personal data protection and intellectual property rights in Singapore.

Important note: The discussion will be conducted on a confidential basis, and we will seek your permission should we intend to use any confidential information you have provided in any public report.

1. Background

1.1. Organisation name

1.2. What is your organisation's main business activity?

1.3. How many people are employed in your Singapore operations?

1.4. How many people / proportion of employees are employed specifically for data and analytics roles globally (if applicable) and in Singapore?

Competition Commission of Singapore

Understanding the Data and Analytics Landscape in Singapore

August 2017

2. Data usage

Understanding the types of data that organisations use

2.1. What types of data are used by your organisation? How are they used? (e.g. customer targeting, market understanding)

3. Data collection and storage

Understanding the types of data that organisations collect/purchase and how they are stored

3.1. If your organisation generates data internally, please answer the following questions:

a. What types of data do you generate? How is it generated? (e.g. provided voluntarily by customers, collected from business operations)

b. At what level of detail is the data generated? (e.g. individual anonymised, by customer profile/location)

c. Are other companies in your industry able to easily generate the same data? Why/why not?

3.2. If your organisation purchases data from external sources, please answer the following questions:

a. What are these external sources and how do they obtain data?

b. What types of data do you purchase, and what is the level of detailed required for such data?

Competition Commission of Singapore

Understanding the Data and Analytics Landscape in Singapore

August 2017

- c. How regularly do you purchase data? Is it the same type of data you purchase every time (e.g. annual updates) or variations of data?

- d. How do you decide what is a “fair” price to pay for the data?

- e. Are there restrictions or conditions applicable when using data from external vendors? (e.g. restrictions on resale, bundling data with other products)

- 3.3. If your organisation does not generate or purchase any data (or only in very limited quantities), please answer the following questions:

- a. Why do you not generate/purchase data, or only do so in limited quantities?

- b. Do you feel the need to start collecting/purchasing data to compete effectively with other businesses in your industry?

- c. What are the main impediments to generating/purchasing more data in your organisation/industry? How can these be overcome?

- 3.4. If your organisation sells data, please answer the following questions:

- a. How do you collect data?

- b. Are organisations able to collect the same data you collect? In other words, what is your organisation’s value-add to these organisations purchasing data?



Competition Commission of Singapore

Understanding the Data and Analytics Landscape in Singapore

August 2017

c. Are there restrictions or conditions applicable when selling data?

3.5. For the data that is collected/purchased, how is it stored? (e.g. network storage devices hosted by the organisation, in servers on the cloud hosted within Singapore, in servers on the cloud hosted outside Singapore)

3.6. What are the key regulations that influence how data is stored, secured and used by your organisation, including local or foreign restrictions?

4. Data analytics

Understanding the role and importance of data analytics and the benefits it brings to organisations

4.1. How much experience does your organisation have with using insights from data and analytics to drive business strategy and growth?

No experience	Under consideration	Being piloted	Recently implemented ¹	Already implemented
---------------	---------------------	---------------	-----------------------------------	---------------------

¹ Recently implemented refers to data and analytics practices implemented in the last 6 months

4.2. If your organisation performs data analytics functions in-house, please answer the following questions:

a. Of all the data collected/procured by your company, how much is processed to gain further value?

< 10%	10 – 40%	40 – 70%	> 70%
-------	----------	----------	-------

b. Do you consider your organisation to be a leader in using data analytics in your sector? Why?

c. Could you provide examples of the types of data analytics your organisation undertakes?

d. How does your organisation currently use data analytics to drive business growth? (e.g. manage risk, attract/retain customers, manage pricing)

e. What are the benefits that the adoption of data and analytics for your company? (e.g. increase revenue/profit, improvements in efficient) Please quantify them, if possible.

Competition Commission of Singapore

Understanding the Data and Analytics Landscape in Singapore

August 2017

- f. Do you think your organisation's use of data analytics brings a competitive advantage? Why?

- 4.3. If your organisation gets external firms to perform data analytics, please answer the following questions:

- a. Who are the main providers of external data analytics services?

- b. How much do you typically spend in a year on such services?

	< \$10K		\$10-50K		\$50-100K		\$100-500K
	\$500K-1m		\$1-5m		\$5-20m		> \$20m

- c. Are there protocols or standards you have to comply with to ensure data privacy is maintained? If so, what are they?

- d. Are there restrictions or conditions applicable when using such services? (e.g. restrictions on sharing analysis, bundling service with other products)

- 4.4. If your organisation only provides data analytics services to other organisations, please answer the following questions:

- a. What types of data analytic services do you provide?

- b. What sectors/industries are your clients typically from?

- c. What are the main impediments to organisations performing more data analytics functions (either in-house or purchased externally)?



Competition Commission of Singapore

Understanding the Data and Analytics Landscape in Singapore

August 2017

4.5. If your organisation does not undertake/purchase any data analytics services (or only in very limited quantities), please answer the following questions:

a. Why do you not undertake/purchase data analytics services, or only do so in limited quantities?

b. Hypothetically, how do you think undertaking/purchasing data analytics services could help in your business? Please elaborate. Do you see the potential benefits exceeding the likely costs?

c. Do you feel the need to start undertaking/purchasing data analytics services to compete effectively with other businesses in your industry?

4.6. What are some of the challenges preventing the greater use of data analytics in your organisation/industry? (e.g. availability of data, nature and form of data collected, availability of talent, legacy systems, non-digital data) How can these be overcome?

Competition Commission of Singapore

Understanding the Data and Analytics Landscape in Singapore

August 2017

5. Sharing of data and data analytics

Understanding the extent of sharing data and analytics, between companies and the incentives/impediments of doing so (i.e. where there are no associated revenues)

5.1. Does your organisation currently share data with other bodies, either other companies or the Government? Why/why not?

5.2. Do you support greater data and analytics sharing between companies in your industry? What are the benefits that can arise from increased sharing? (e.g. increase in revenue/profits) Please quantify the benefits if possible.

5.3. Are you more likely to share data with companies outside of your industry than within your industry? Why?

5.4. If data and analytics are being shared, what types of data are shared and what are the arrangements are in place to facilitate this? (e.g. mutual agreement between companies by contract, collected by independent third party)

5.5. If data and analytics are being shared, how is compliance with data protection laws ensured?

5.6. What are the main challenges to greater sharing of data? (e.g. companies' unwillingness to share proprietary data, unstructured data that cannot easily be processed by different companies) How can these be overcome?

5.7. What initiatives (either Government or industry-level) would help promote greater data sharing?

Competition Commission of Singapore

Understanding the Data and Analytics Landscape in Singapore

August 2017

6. Monetisation of data and analytics

Understanding how data and analytics are being monetised (i.e. the generation of revenues from business practices around data and analytics)

6.1. How is revenue being generated from data and data analytics use your organisation?

Method	Description
Business optimisation	
Sell raw data	
Sell processed data (after data analytics)	
Sell data analytics services	
Others	

6.2. For each of the methods mentioned, how much is typically monetised in a year?

	< \$10K		\$10-50K		\$50-100K		\$100-500K
	\$500K-1m		\$1-5m		\$5-20m		> \$20m

6.3. Are there third party organisations that capture and monetise data and analytics in your industry? Who are they and how do they do that?

6.4. Are there laws and regulations that govern the monetisation of data? How does your organisation ensure compliance with these?

6.5. What are the trends to data monetisation? Which of the methods listed above will have the highest growth potential?

6.6. What are the main impediments to wider monetisation of data and analytics? How can these be overcome?



Competition Commission of Singapore

Understanding the Data and Analytics Landscape in Singapore

August 2017

7. Outlook and trends

Understanding the growth and future scenarios of data and analytics

7.1. How has the collection of data and use of data analytics transform your industry, including the operations, business models and performance?

7.2. How do you see the collection of data and use of data analytics changing in your industry in the future?

7.3. Would the possession of data and analytics skills become critical to operations in your industry?

7.4. What are the main challenges to your organisation from the growing importance of data and analytics in your industry?

7.5. What are the main opportunities to your organisation from the growing importance of data and analytics in your industry?

8. Other comments

8.1. Do you have any other comments on the data and analytics landscape in Singapore to add?